# Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
*Department of Psychology, University of Michigan*
*Ann Arbor, Michigan*

DAVID A. GRANT, Associate Editor
*University of Wisconsin*

WILLIAM K. ESTES, Associate Editor
*Indiana University*

Consulting Editors

NORMAN H. ANDERSON, *University of California, Los Angeles*
E. JAMES ARCHER, *University of Wisconsin*
FRED ATTNEAVE, III, *University of Oregon*
CLETUS J. BURKE, *Indiana University*
JAMES DEESE, *Johns Hopkins University*
PAUL M. FITTS, *University of Michigan*
FRANK A. GELDARD, *University of Virginia*
JAMES J. GIBSON, *Cornell University*
CLARENCE H. GRAHAM, *Columbia University*
HAROLD W. HAKE, *University of Illinois*
ARTHUR L. IRION, *Tulane University*
HOWARD H. KENDLER, *New York University*
DONALD B. LINDSLEY, *University of California, Los Angeles*
KENNETH MACCORQUODALE, *University of Minnesota*
QUINN MCNEMAR, *Stanford University*
L. STARLING REID, *University of Virginia*
KENNETH W. SPENCE, *State University of Iowa*
DELOS D. WICKENS, *Ohio State University*

## CONTENTS

Effects of Increments of Reinforcement in Human Probability Learning: M. W. SHELLY II ... 345
Identification of Visual Correlational Scatterplots: I. POLLACK ... 351
Task Difficulty and Drive in Verbal Learning: R. R. STANDISH AND R. A. CHAMPION ... 361
Association Value and Familiarity in Serial Verbal Learning: R. H. LINDLEY ... 366
Effects of Probability of Misinformation and Number of Irrelevant Dimensions upon Concept Identification: V. PISHKIN ... 371
Eyelid Conditioning Performance under Partial Reinforcement as a Function of UCS Intensity: L. E. ROSS AND K. W. SPENCE ... 379
Effect of Triplet and Quadruplicate Location in Verbal Maze Patterns upon Serial Position Errors: G. NAMIKAS, C. P. THOMPSON, AND W. J. BROGDEN ... 383
Effect of Concurrent Responses on the Evocation and Generalization of the Conditioned Eyeblink: G. R. GRICE AND J. D. DAVIS ... 391
Roles of Association Value and Syllable Familiarization in Verbal Discrimination Learning: W. N. RUNQUIST AND M. FREEMAN 396
Effect of Runway Size and Drive Strength on Acquisition and Extinction: D. J. LEWIS AND J. W. COTTON 402
Learning and Retention of Paired Adjectives as a Function of Percentage Occurrence of Response Members: R. W. SCHULZ AND W. N. RUNQUIST ... 409
A Test for Interaction of Delay of Knowledge of Results and Two Types of Interpolated Activity: E. A. BILODEAU AND F. J. RYAN ... 414
Effects of Situational Anxiety at Different Stages of Practice: J. L. BARDACH ... 420
Role of Postural Experiences in Proprioceptive Perception of Verticality: R. G. PEARSON AND G. T. HAUTY 425
Supplementary Report: The Yerkes-Dodson Law and Shift in Task Difficulty: V. H. DENENBERG AND G. C. KARAS ... 429
Supplementary Report: Further Factors Affecting the Probability of Changing Responses: M. W. SHELLY II 430
Supplementary Report: Familiarity and Frequency: C. E. NOBLE ... 432
Supplementary Report: The Effect of Verbal Reinforcement Combinations on Learning in Children: C. CURRY ... 434

American Psychological Association

Vol. 59, No. 6 June 1960

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $8.00, or $16.00 annually. Single copies are $1.50. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This Journal publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects, except when the latter are specifically oriented toward the extension of behavior theory. Experimental psychometric studies and studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology.

Normally, articles exceeding 20 printed pages cannot be accepted. Within this limit, the piecemeal experiment-by-experiment reporting of psychological research is discouraged, and the reporting of the data and conclusions of substantial segments of research efforts is encouraged. Specifically, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) *must* be presented in a single article. Special provision is made for the publication of Supplementary Reports and Replication Reports in articles of not more than 1.5 printed pages (see Editor's Note, this Journal, 1957, **53**, 1–2).

Address all articles submitted for publication to the Editor, Arthur W. Melton, Department of Psychology, University of Michigan, Ann Arbor, Michigan. **Manuscripts *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc., as described in the *Publication Manual of the American Psychological Association.*** When in doubt about practices of the Journal, authors should examine a recent issue.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1960 the publication cost is $30 per printed page.

To save printing costs, supplementary material to articles in this Journal are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

---

ARTHUR C. HOFFMAN, Managing Editor
HELEN ORR, Promotion Manager
BEATRICE B. HAMLIN, Editorial Assistant

Second-class postage paid at Lancaster, Pa.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1960 by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 59, NO. 6 JUNE 1960

# EFFECTS OF INCREMENTS OF REINFORCEMENT IN HUMAN PROBABILITY LEARNING[1]

MAYNARD W. SHELLY II

*Office of Naval Research*

There are indications in the literature that even though relative frequencies of reinforcement may have proportionate effects on response frequencies asympototically, they may not have such proportionate effects at preasymptotic levels (e.g., Grant, Hake, & Hornseth, 1951). Such indications can be observed when trial or block number replaces reinforcement conditions as the parameter of the family of performance curves, and reinforcement conditions replace trial or block number on the abscissa in the usual graphs of results. From similar plots there are also faint evidences (e.g., Neimark, 1956) that any lack of preasymptotic proportionality may be affected by the type of reinforcement, i.e., whether it is contingent or noncontingent. This experiment was conducted to determine how the relative frequency of a response with some fixed probability of reinforcement is affected by variations in the probability of reinforcement of alternative responses.

To study the effects of variations in the probability of making alternative responses, this investigation employed a situation approximating the usual probability learning conditions. Three alternatives were used, and *S*s were asked to predict which of the three alternatives would be correct on a given trial. None, one, two, or three alternatives could be correct on any one trial. Because of the indications that contingent and noncontingent reinforcement might produce different effects on another alternative, even for a constant probability of reinforcement, both conditions of reinforcement were used.

This study uses three, rather than two, response alternatives for two principal reasons. The first is that with three alternatives it is possible to have more sizable variations in the

[1] This research was conducted while the author was at the University of Wisconsin. He wishes to express his appreciation to the Graduate School of the University of Wisconsin for providing funds for the experimental apparatus employed, and to the Social Science Research Council for providing the author with a Research Training Fellowship for the period during which the experiment was completed. He also wishes to thank the Office of Naval Research for providing the author with time, and the Department of Psychology, Northwestern University for the facilities to complete the final drafts of this paper.

sum of the probabilities of reinforcement on two of the three alternatives and still have the probability of reinforcement on the third a maximum. Secondly, it is only when three or more alternatives are used that contingent and noncontingent reinforcement are fully differentiated.

## METHOD

*Subjects.*—The *S*s were 170 students (approximately evenly divided between women and men) from elementary psychology courses at the University of Wisconsin.

*Apparatus.*—The apparatus consisted of a booth with a flat black interior in which *S* sat, and equipment outside the booth which regulated the experimental conditions. In the booth an aluminum chassis box (8 in. × 17 in.) was mounted on a sloping table. It contained 16 four-pole snap switches, above each of which was mounted a 6-v. green-domed lamp. For the conditions of this study, only the three switches and their associated lamps were uncovered. A yellow-domed 6-v. lamp was also mounted in front of *S*. This informed him that a new trial had begun. When it went off, *S* was to make a choice of one of the three switches before him.

The equipment outside of the booth consisted largely of a timing motor, Western Union tape transmitters, and lamps on a panel which permitted *E* to record *S*'s responses. There were also a number of rotary switches which permitted *E* to change the correspondence between channels of the Western Union tape transmitters and *S*'s response alternatives.

*Design.*—The experiment consisted of 17 conditions, 10 different *S*s serving in each condition. In all 17 conditions at least one alternative would lead to a correct prediction on 75% of the trials if always chosen. This alternative was combined with two other response alternatives. The reinforcement probabilities on these other two alternatives were combinations of 0, .25, .50, and .75. Random binary sequences were established independently for each alternative, so that even when the probabilities summed to one, there were trials on which no switch was correct and trials on which one or more switches would be correct.

The two types of reinforcement were contingent and noncontingent reinforcement. When *S* served under contingent reinforcement, he was only informed as to whether or not he had made a correct prediction. If he had been correct, the lamp above the alternative he had chosen came on. If he had been incorrect in his prediction, no lamp came on. When noncontingent reinforcement was used, a lamp came on above each correct alternative (if there were any) irrespective of whether *S* had made a correct or incorrect prediction.

TABLE 1

COMBINATIONS OF REINFORCEMENT PERCENTAGES USED IN THE EXPERIMENTAL CONDITIONS

| | | |
|---|---|---|
| 75-0-0[a] | | |
| 75-0-25 | 75-25-25 | |
| 75-0-50 | 75-25-50 | 75-50-50 |
| 75-0-75 | 75-25-75 | 75-50-75 |

Note.—The numbers in the table represent the percentages of times the first, second, and third alternatives, respectively, would be reinforced if constantly chosen.

[a] This condition was not completed under noncontingent reinforcement because *S*s were responding perfectly after only a few trials, and a factor of boredom would have been too strong a confounding influence.

The design is presented in Table 1. Here the first "75" refers to the presence of an alternative which was reinforced 75% of the time. The two numbers following the initial "75" give the percentages of reinforcements which would be received were either of the other alternatives constantly chosen. The effects of any consistent position preferences among *S*s were eliminated by permuting the positions of the various alternatives among *S*s.

*Procedure.*—The *S*s in this experiment served individually. When they reported to the laboratory, they were assigned to one of the 17 conditions on a random basis with the restriction that no more than 10 *S*s be assigned to any one condition. Upon being seated in the experimental booth, all *S*s were given the same standard set of instructions (except for modifications appropriate to the conditions of reinforcement). These instructions told *S*s that the experiment was designed to study their abilities to predict future events, and that the object of the experiment was to make as many correct predictions as possible. Also included were instructions regarding the mechanics of the experiment. It was strongly suggested that they were to make one choice on each trial, and never more than one. A brief demonstration was then given.

Each *S* then made a total of 320 predictions. Each trial was of 4-sec. duration.

## Results

In Fig. 1 are presented the performance curves for the 75% alternative for the conditions of the present experiment. At the termination of each curve are two numbers separated by a hyphen. These numbers represent the percentages of times that the two alternatives not plotted for that group could lead to reinforcement in 100 trials. The number in parentheses associated with each curve refers to the asymptote which would be approached were *S* responding proportionately (in the following sense) to the percentage of reinforcement. A response will be considered to be proportionate if the proportion of times an alternative is chosen is equal to the ratio of the number of times that alternative can be reinforced to the sum of the number of times all alternatives can be reinforced. For the 75-25-25 condition the ratio for the 75% alternative would be $75/(75 + 25 + 25)$. As can be seen, the departures from such proportionality are greatest when the denominator of the above ratio is smallest.

In the introduction it was stated that there are suggestions that increments in amount of reinforcement assigned to other alternatives do not always have proportionate decremental effects on the response alternative being analyzed. Such a suggestion seems to be further confirmed by the performance curves in Fig. 1. The *D*s in Fig. 1 are the distances between curves *when* the totals of the two numbers identifying the curves involved differ by 25. (To keep the graphs relatively simple only two of the possible *D*s are shown.) To consider an example, take Part a of Fig. 1; here $25 + 0$ and $50 + 0$ differ by 25 and an arrow is drawn between the two curves (for the last block of trials only) and labeled by *D*. It can be observed that the *D* distances are not always proportionate to the differences between the numbers included in the parentheses. In an attempt to analyze such relationships more formally, the following type of decision procedure was used.

We can specify the percentages of reinforcement for two alternatives and let the third vary over the values 0, 25, 50, and 75, which stand for percentages of reinforcement. Denote the third alternative, which can vary, by X. Then, using { } to denote a set, {75-0-X} is, as an example, the set of conditions {75-0-0, 75-0-25, 75-0-50, 75-0-75}. Such sets of conditions will be referred to as *C sets*. It will be observed that when the members of a C set are written in the above ordering, the percentage of reinforcement for the last alternative increases by steps of 25; i.e., the X values increase by steps of 25. By using the *D* distances as ordinate values, and X values on the abscissa for a fixed C set of conditions we get a family of curves, the parameter of which is trial block number. For each of the two types of reinforcement there are four C sets: {75-0-X}, {75-25-S}, {75-50-X}, and {75-75-X}. Clearly, the conditions of the experiment usually do not belong to only one C set.

Because X can assume four values, there are three *D* distances possible, and hence each curve, for a given block of trials, based upon a given C set, contains three points, the first of which is at an X value of 25. Each of these three points can be interpreted as the decrease in number of times the first alternative (the alternative which in every condition could lead to a reinforcement 75% of the time) is responded to as a

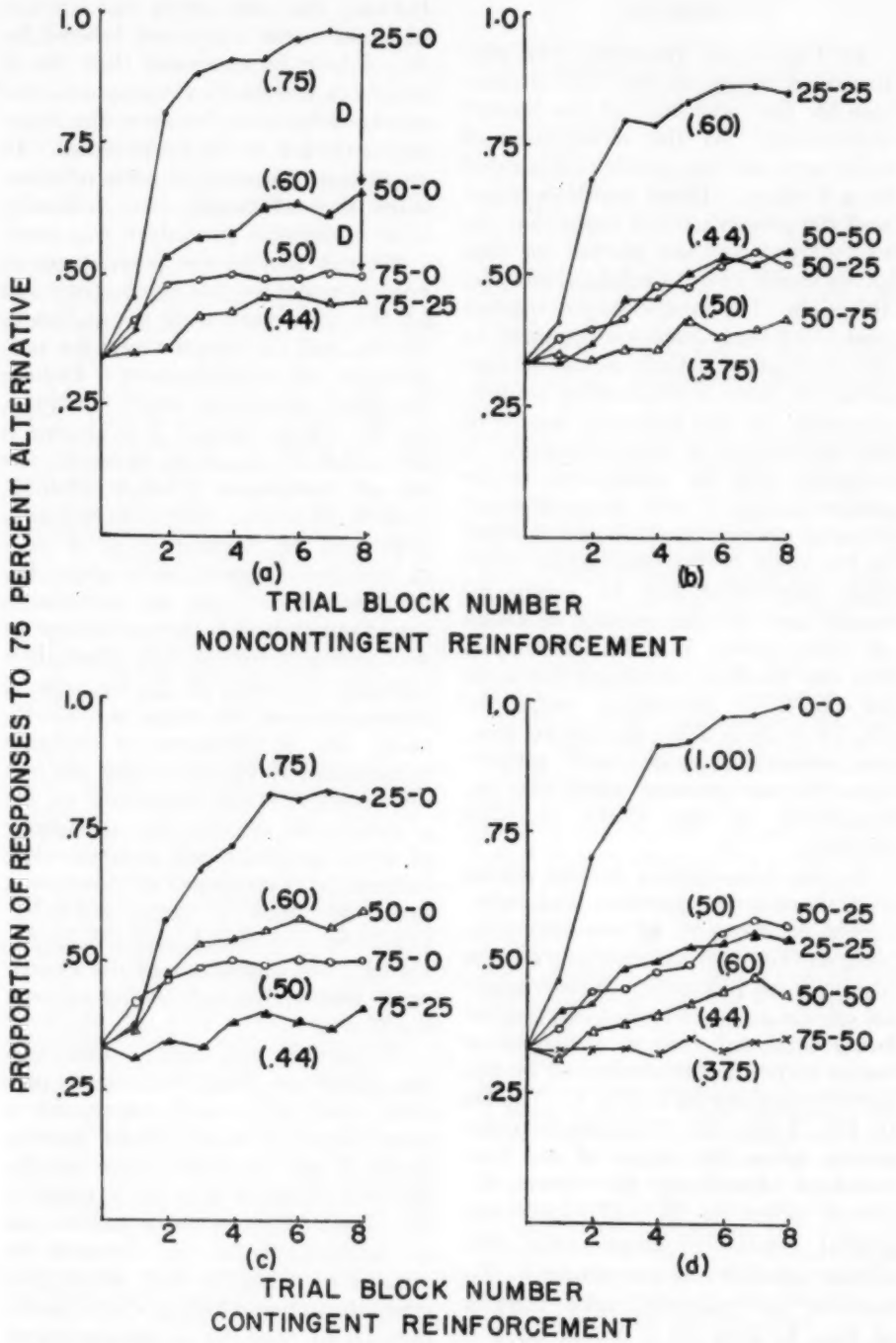



FIG. 1. Proportion of responses to the 75% alternative in all conditions for Trial Blocks 2, 4, 6, and 8. The division of performance curves between Parts a and b, and between Parts c and d is simply to make them more easily readable.

result of an increase in the X value by 25, e.g., from 0 to 25, or from 25 to 50. It can easily be computed, using the definition of proportionality in this paper, that the maximum point on each of these curves should be the first point, i.e., the $D$ distances between 75-0-0 and 75-0-25 for example.

Families of curves based upon the C sets were plotted for the second, fourth, sixth, and eighth block of 40 trials. Sequences of one or two hypotheses were then tested. The first was that twice the maximum point on each curve was greater than the sum of the other two. This simply tested the statistical hypothesis that there was a statistically significant maximum point. The procedures used in this and the following hypothesis are those outlined in Grant (1956). For example, if the second point were the greatest, the hypothesis tested would be:

$$-(\mu_0 - \mu_{25}) + 2(\mu_{25} - \mu_{50}) - (\mu_{50} - \mu_{75}) = 0 \quad [1]$$

where $\mu_0$, for example, is the mean for that condition of the C set for which X = 0.

If this hypothesis were not rejected, then the hypothesis that the sum of the two greatest points was greater than twice the remaining point would be tested. For example, if this happened to be the first two versus the third point, the hypothesis tested would be:

$$(\mu_0 - \mu_{25}) + (\mu_{25} - \mu_{50}) - 2(\mu_{50} - \mu_{75}) = 0 \quad [2]$$

where $\mu_0$, $\mu_{25}$, etc. have the same meanings as in [1]. If either one of the hypotheses were rejected, then a statistical maximum would be reached at one, or at two points. These inferred statistical maxima, max($D$), are given in Table 2. It is very evident that max($D$), when it occurs, does not always occur at an X value of 25. Indeed, 62.5% of them occur at an X value of 50.

TABLE 2

VALUES OF X FOR WHICH MAX($D$) OCCURS

| C set | Trial Block Number | | | |
|---|---|---|---|---|
| | 2 | 4 | 6 | 8 |
| **Noncontingent Conditions** | | | | |
| {75-0-X}[a] | 50 | 50 | 50 | 50 |
| {75-25-X} | 50 | 50 | 50 | 50 |
| {75-50-X} | 25 | — | — | (25, 75) |
| {75-75-X}[b] | — | — | — | — |
| **Contingent Conditions** | | | | |
| {75-0-X} | — | (25, 50) | — | 50 |
| {75-25-X} | — | — | (25, 75) | (25, 75) |
| {75-50-X} | — | — | — | — |
| {75-75-X}[b] | 25 | — | — | — |

[a] For 75-0-0 the value of 1.00 was used for the proportion of responses to the 75% alternative and the statistical procedures adjusted accordingly.

[b] For 75-75-75 the value of .333 was used for the proportion of responses to the 75% alternative and the statistical procedures adjusted accordingly for use of a nonempirical value.

## DISCUSSION

Obviously the more complete implications of this study can only be arrived at through an extended mathematical analysis. There are, however, several things about the findings which can be emphasized. First, the max($D$) values do not usually occur at an X value of 25. This implies that the probabilities of reinforcement are not being responded to proportionately. This result agrees with the findings of Gardner (1957) and Cotton and Rechtschaffen (1958). Secondly, there is considerable consistency in the X values at which the max($D$)s occur for different blocks of trials. This implies some consistency in the relative learning rates. Thirdly, 62.5% of the max($D$)s occurred at an X value of 50. That is, the greatest decrease in responding to the 75% alternative occurred when the proba-

bility of reinforcement on the variable alternative in a C set reached maximal uncertainty, i.e., changed from a probability of .25 to a probability of .5. It is possibly this last finding which is most suggestive for future research.

## Summary

An experiment was conducted to determine how the relative frequency of a response with some fixed probability of reinforcement is affected by variations in the probability of reinforcement of alternative responses. The experimental situation approximated the usual probability learning conditions. The *Ss* were asked to predict which of three events would occur on any one trial. One of the alternative responses had a probability of reinforcement of .75 for all *Ss*. The second and third responses had probabilities of reinforcement varying from 0 to .75 in steps of .25 for the various experimental groups. The results were found not to be in agreement with a prediction based upon a generalization of the meaning of proportionality for two alternative responses. Instead it was found that, when the reinforcement value of the first two alternative responses was kept constant, the greatest decrement in responding to the first (the .75 alternative) occurred when the probability of reinforcement on the third alternative increased from .25 to .5.

## References

Cotton, J. W., & Rechtschaffen, A. Replication report: Two- and three-choice verbal-conditioning phenomena. *J. exp. Psychol.*, 1958, **56**, 96.

Gardner, R. A. Probability learning with two and three choices. *Amer. J. Psychol.*, 1957, **70**, 174–185.

Grant, D. A. Analysis of variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53**, 141–154.

Grant, D. A., Hake, H. W., & Hornseth, J. P. Acquisition and extinction of a verbal conditioned response with differing percentages of reinforcement. *J. exp. Psychol.*, 1951, **42**, 1–5.

Neimark, E. D. Effects of type of nonreinforcement and number of alternative responses in two verbal conditioning situations. *J. exp. Psychol.*, 1956, **52**, 209–220.

(Received June 26, 1959)

*Journal of Experimental Psychology*
Vol. 59, No. 6, 1960

# IDENTIFICATION OF VISUAL CORRELATIONAL SCATTERPLOTS [1]

IRWIN POLLACK

*Operational Applications Laboratory, Air Force Cambridge Research Center*

A useful tool for the examination of interaural phenomena is the description of auditory signals at the two ears in terms of their correlation with each other (Pollack & Trittipoe, 1959; Zerlin, 1959). The aim of the present paper is to describe initial studies with a visual correlational display. This display, we believe, provides a useful tool for the examination of dynamic shape perception and the behavioral testing of statistical hypotheses.

The student of psychological statistics is familiar with the correlational scatterplot. If two variables, $A$ and $B$, strongly share a single common factor, $C$, the correlational scatterplot of $A$ versus $B$ will be confined to a small area about the diagonal. If the two variables weakly share the common factor, $C$, the correlational scatterplot may be diffuse and may cover a large extent of the entire $A$ versus $B$ area. In terms of components of variance, the correlation (squared) is equal to the ratio of the variance common to each of two variables to the total variance (common variance plus independent variance associated with each of the two variables).

Licklider and Dzendolet (1948, Fig. 3, p. 123) have demonstrated a simple procedure for producing correlational distributions. Three independent random noise generators, $A$, $B$, and $C$, are employed. If the output of Noise Generator $A$ is sent to the $x$ coordinate of an oscilloscope and the output of Independent Noise Generator $B$ is sent to the $y$ coordinate, a diffuse correlational plot with a circular envelope will be obtained. Since $A$ and $B$ are independent random distributions, a correlation of zero will be obtained. If, on the other hand, the output of Noise Generator $C$ alone is sent to both coordinates of the oscilloscope, a thin line at 45° will be traced if the $x$ and $y$ coordinate gains are equal. Since a single distribution is sent to both coordinates, a correlation of 1.0 will be obtained.

If $C$ is mixed with $A$ and sent to the $x$ coordinate; and if $C$ is mixed with $B$ and sent to the $y$ coordinate, intermediate correlations can also be obtained. The greater the relative contribution of the common source, $C$, the higher will be the correlation. The greater the relative contribution of the independent sources, the smaller will be the correlation. And, in general, the correlation (squared) will be equal to the ratio of the common noise power, $C$, over the total noise power $(C + A)$ or $(C + B)$. Licklider (1948, Fig. 4, p. 155) presents oscillographic pictures traced out by a rapidly moving electron stream for a wide range of correlations. Licklider and Dzendolet (1948, Fig. 2, p. 122) also present oscillographic

[1] This report is AFCRC TR 59–14 of the Air Force Cambridge Research Center. This research supports Project 7682, Information Processing by the Human Operator, of the Air Research and Development Command Program in Human Performance. The writer wishes to acknowledge the able assistance of William Trittipoe in all phases of the testing program, and John Schjelderup with respect to the experimental equipment.

pictures obtained by discrete sampling of correlational distributions.[2]

## Method

The task of *S* was to identify one of two correlational distributions presented on an oscilloscope. In terms of three independent noise generators, a reference correlation was obtained by routing $(C + A)$ to one coordinate of the oscilloscope and by routing $(C + B)$ to the other coordinate. In all tests, $A = B$; and $(C + A) = (C + B)$. A second, higher, correlation was obtained by mixing a higher level of the common source, $C'$, with the two independent noise sources. The task of *S* was to identify whether the correlation was produced by $(C + A)$ and $(C + B)$ or by $(C' + A)$ and $(C' + B)$. In terms of the verbalization of *S*'s task, he was instructed to press his voting lever "down" for the "stubbier" of the two scatterplots and to press his voting lever "up" for the "thinner" of the two scatterplots. It is important to note that the power sum of $(C' + A)$ was equated to $(C + A)$ in order that distributional differences, rather than differences in amplitude, would serve as the relevant cue for discrimination.

Factors which facilitated rapid learning of the task were: (*a*) immediate confirmation of correctness of response was provided; (*b*) a consistent direction of movement of *S*'s voting lever was assigned to the "thinner" of the two correlations; (*c*) *S*s worked in a group and the score of each *S* was announced to the entire group after each set of 30 or 60 presentations.

The oscilloscopes employed were Dumont Model 304-A, one with a fast blue P-11 phosphor, one with a medium persistence green P-1 phosphor. Since no consistent difference was observed between the results yielded by the two tubes, only the average results are reported. Approximately three-quarters of the 5-in. diameter was employed for the display. The noise generators were Grason-Stadler Model 455 B, set at 20 kcps. The noise spectrum was uniform between 100 and 6800 cps after being passed through a band-pass filter with an attenuation characteristic of 18 db/octave. Low-pass filtering was obtained by sharp *m*-derived filters, Gertsch Model BP-1. A variable phase-shift circuit was employed in order to sharpen the scope image for high noise frequencies. The interval between successive correlation displays was 4 sec.

Discrete samplings from the correlational distributions were obtained in the following manner. The two correlational distributions were first established, and the intensity control of the oscilloscope was reduced so that a picture was not visible. The intensity axis, or *z* axis, of the oscilloscope was then pulsed to yield short discrete samples from the correlational distribution. The pulse duration was 6 μsec., obtained from a Measurements Corporation 79-B pulse generator.

In all, eight separate experimental series (to be described separately) were carried out. In each series, a reference correlation $(C + A)$ was established, and a wide range of variable correlations $(C' + A)$ was explored in order that the point of 75% correct two-choice identification could be established. In all tests except Series IV, the average seating distance from the scopes was about 18 in. In Series IV, the distance was 54 in. A nonpulsed display was employed in Series I–V; a sampled discrete-pulsed display was employed in Series VI–VIII.

Series I consisted of four *S*s, all laboratory personnel experienced in psychophysical testing. Series II–V employed six *S*s, five female college undergraduates and one of the *E*s. Prior to the testing series, the undergraduates had not participated in psychophysical studies. Series VI–VIII employed six *S*s, four female college undergraduates and both *E*s, all experienced in psychophysical studies. Each presented datum point for for Series II–VIII represents the interpolation from a psychophysical function based on about 1800 observations—6 *S*s, 4–6 variable correlations, and 60 observations by each *S* for each variable correlation, except Series V the corresponding number of observations was about 5600.

## Results

### *Series I: Preliminary Tests*

The purpose of Series I was to compare the results of visual tests with the results of auditory tests that were carried out with the same noise distributions (Pollack & Trittipoe,

[2] Even the excellent still photographs in the two references fail to convey the dazzling dancing parade of correlation patterns available for study. To the serious reader interested in dynamic displays with information of well defined statistical properties, a lash-up of the noise generators and oscilloscope is recommended to taste the full richness and flavor of the available materials.

1959). In the auditory tests, $(C + A)$ or $(C' + A)$ was presented to one earphone and $(C + B)$ or $(C' + B)$ was presented to the other. The abscissa of Fig. 1 is the reference correlation (squared); the ordinate is the difference between the reference correlation (squared) and the variable correlation (squared) required for 60, 75, and 90% correct identification. The visual tests are represented by solid lines. The thin dashed curve represents 60% correct identification for the auditory tests.

Two main points are suggested by Fig. 1: (*a*) In light of the vastly different tasks with the auditory and visual displays, the form of the correlation identification function for auditory and visual tests is strikingly similar. The 60% correct auditory contour nearly overlaps with the 90% correct visual contour. (*b*) The three visual correlation identification functions exhibit roughly the same general form. Because the two-choice identification function is nearly linear about its midpoint, the 75% correct identification level was selected to summarize the results of all of the other tests.

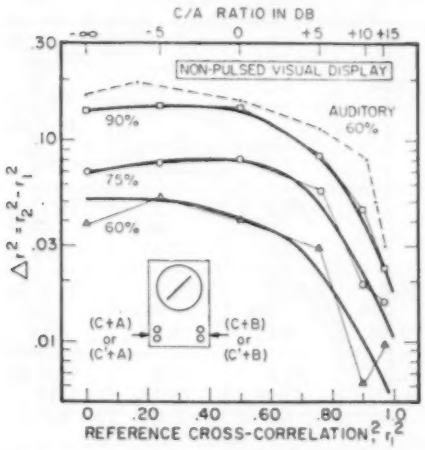



FIG. 1. Comparison of auditory and visual correlational displays. The abscissa is the reference correlation (squared). The ordinate is the difference between the required correlation (squared) for 60, 75, and 90% correct identification and reference correlation (squared). The ordinate is scaled logarithmically. The dashed line represents the results of auditory tests for 60% correct identification. Tests of Series I.

*Series II: Effect of Duration*

The results of Series II are presented in Table 1. A given testing condition is defined by the display duration (across the top) and the reference correlation (along the left side). For each condition, three entries are presented: (*a*) the $C'/A$ ratio in decibels required for 75% correct identification when paired with the reference correlation on the left; (*b*) the equivalent correlation (squared); and (*c*) the equivalent Fisher $z$ transformation score. The three corresponding entries are also presented for the reference conditions on the left.

The effect of duration is not marked except for the conditions of high reference correlation where extremely large differences were needed to obtain 75% correct identification. (The 75% correct identification level could not be obtained for the 10-msec. display for the two highest reference correlations within the ranges of conditions available.)

The ratio between $C$ and $C'$ (in terms of the decibel difference required for 75% correct identification) furnishes one measure of relative sensitivity. In terms of this measure, maximum sensitivity is obtained near a reference correlation (squared) of about 0.5 with reduced sensitivity for lower and higher reference correlations.

Another measure of sensitivity is the difference between the correlations (squared) required for 75% correct identification. The results

TABLE 1

IDENTIFICATION OF TWO VISUAL CORRELATIONS AS A FUNCTION OF THE REFERENCE CORRELATION AND DISPLAY DURATION

(For each condition, three entries are presented: (*a*) the *C'/A* ratio in decibels which was correctly identified 75% of the trials from the reference *C/A* ratio listed on the left; (*b*) the equivalent correlation coefficient (squared); and, (*c*) the equivalent Fisher *z*-score transformation.)

| Reference Correlation | Measure | Duration in Milliseconds | | | | |
|---|---|---|---|---|---|---|
| | | 10 | 32 | 100 | 316 | 1000 |
| C/A $-\infty$ | C'/A | −4.3 | −6.3 | −7.7 | −9.4 | −10.0 db |
| $r^2$ .00 | $r^2$ | .27 | .19 | .15 | .10 | .09 |
| *z* .00 | *z* | .58 | .47 | .41 | .32 | .31 |
| C/A −20 db | C'/A | −3.7 | −7.9 | −8.8 | −9.2 | −10.0 db |
| $r^2$ .01 | $r^2$ | .30 | .14 | .12 | .11 | .09 |
| *z* .10 | *z* | .61 | .39 | .36 | .35 | .31 |
| C/A −10 db | C'/A | −4.0 | −4.8 | −6.1 | −7.0 | −7.0 db |
| $r^2$ .09 | $r^2$ | .28 | .25 | .20 | .17 | .17 |
| *z* .31 | *z* | .59 | .55 | .48 | .44 | .44 |
| C/A −5 db | C'/A | −0.8 | −1.3 | −2.1 | −2.8 | −3.4 db |
| $r^2$ .24 | $r^2$ | .47 | .42 | .38 | .34 | .32 |
| *z* .54 | *z* | .84 | .77 | .72 | .67 | .64 |
| C/A 0 db | C'/A | 3.3 | 2.6 | 1.7 | | 1.3 db |
| $r^2$ .5 | $r^2$ | .67 | .64 | .60 | | .57 |
| *z* .88 | *z* | 1.15 | 1.10 | 1.03 | | .99 |
| C/A +5 db | C'/A | 7.8 | 7.0 | 6.6 | | 6.0 db |
| $r^2$ .76 | $r^2$ | .86 | .83 | .82 | | .80 |
| *z* 1.34 | *z* | 1.64 | 1.54 | 1.50 | | 1.44 |
| C/A +10 db | C'/A | 14.3 | | 12.0 | | 11.7 db |
| $r^2$ .91 | $r^2$ | .96 | | .94 | | .92 |
| *z* 1.87 | *z* | 2.30 | | 2.09 | | 1.94 |
| C/A +15 db | C'/A | 37.0 | | 18.7 | 17.2 | 16.7 db |
| $r^2$ .97 | $r^2$ | .(3)8[a] | | .983 | .981 | .979 |
| *z* 2.43 | *z* | 5.0 | | 2.73 | 2.65 | 2.62 |
| C/A +20 db | C'/A | — | 25.8 | 21.9 | 21.7 | 21.2 db |
| $r^2$ .990 | $r^2$ | — | .997 | .994 | .993 | .992 |
| *z* 2.95 | *z* | — | 3.60 | 3.25 | 3.17 | 3.13 |
| C/A +25 db | C'/A | — | 41.0 | 32.3 | 32.0 | 31.6 db |
| $r^2$ .997 | $r^2$ | — | .(4)2[b] | .(3)4[a] | .(3)3[a] | .(3)2[a] |
| *z* 3.58 | *z* | — | 5.8 | 4.4 | 4.4 | 4.3 |

[a] .(3)8 .9998.
[b] .(4)8 .99998.

are presented in Fig. 2 in terms of the $\Delta r^2$ measure. The lower abscissa is the *C/A* ratio in decibels; the upper abscissa is the equivalent correlation (squared). The parameter on the curves is the display duration. In terms of this measure, sensitivity changes little until a reference correlation (squared) of about 0.5 with rapid improvement, thereafter, for higher reference correlations.

Another measure of sensitivity is

the difference between the correlations in terms of a metric which normalizes the sampling variance of the correlation coefficient. If successive samples are, for example, obtained from a population correlational distribution with a correlation of near 1.0, the individual sample correlations will all be near 1.0. However, if successive samples are obtained from a population correlational distribution with a correlation of nearly zero, the sample correlations may vary over a wide range. That is, the sampling error is much wider for correlations near zero than for correlations near 1.0. The Fisher $z$ transformation attempts to equalize the sampling variance for all correlations. In terms of the $\Delta z$ measure, over most of the reference correlations, the range of scores is considerably reduced compared with the range of scores in terms of $\Delta r^2$ or $\Delta r$. We tentatively interpret this result to indicate that $S$'s task is to sample the correlational distributions and to make a test between alternative statistical hypotheses. That performance is also reduced with shorter durations also lends support to the notion that $S$ is acting as a tester of alternative statistical hypotheses.

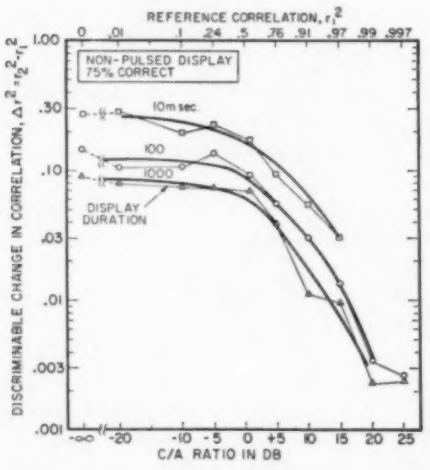



FIG. 2. Identification of two visual correlations. The upper abscissa is the reference correlation (squared); the lower abscissa is the equivalent $C/A$ ratio in decibels. The ordinate is the difference between the required correlation (squared) for 75% correct identification and the reference correlation (squared). The ordinate is scaled logarithmically. The parameter is the display duration in milliseconds. Tests of Series II.

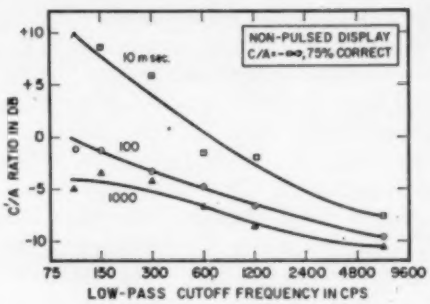



FIG. 3. Effect of low-pass cutoff frequency upon identification of two visual correlations. The abscissa is the low-pass cutoff frequency in cycles/second. The ordinate is the $C'/A$ ratio in decibels required for 75% correct identification from a $C/A$ ratio of $-\infty$. The parameter is the display duration in milliseconds. Tests of Series III.

### *Series III: Effect of Frequency Range*

The effect of low-pass filtering of the noise distributions was examined for a zero reference correlation for three durations: 10, 100, and 1000 msec.[3] In terms of the stimulus to $S$, the effect of low-pass filtering is to thin out the density of the correlational distribution upon the oscilloscope. This operation may be alternatively considered from a sampling viewpoint as a reduction of the sample size available for inspection. In terms of the testing of alternative statistical hypotheses, we would there-

[3] The equivalence of the correlation (squared) with the ratio of the noise powers may be questioned for restricted bands of noise.

fore expect poorer discrimination with successively deeper cuts in the noise frequency distribution.

The results of the filtering tests are presented in Fig. 3. The abscissa is the low-pass cutoff frequency. The ordinate is the $C'/A$ ratio necessary to achieve 75% correct identification. The parameter is the duration of the correlational display. The reference correlation was zero.

Larger $C'/A$ ratios are required as the frequency range is reduced. The effect is most marked with the display of shortest duration.

*Series IV: Effect of Size*

If a given correlation is established, a change in size of one or both axes will leave the correlation invariant. If *S*'s task is primarily that of identifying differences in correlation, we would expect little change in identification with changes in size (until acuity considerations become important).

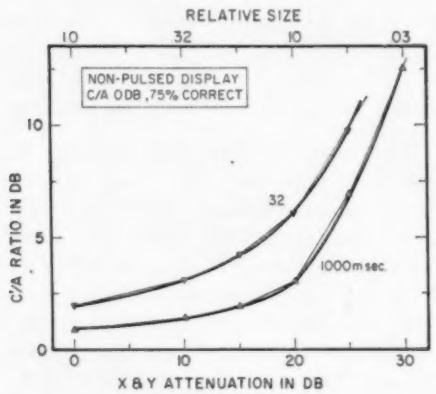



FIG. 4. Effect of size of scatterplot upon identification of two correlations. The lower abscissa is the attenuation in decibels applied to both coordinates of the oscilloscope. The upper abscissa is the relative linear size of the resulting scatterplots. The ordinate is the $C'/A$ ratio in decibels required for 75% correct identification from a reference $C/A$ ratio of 0 db. The parameter is the display duration in milliseconds. Tests of Series IV.

The effect of the size of the scatterplot is presented in Fig. 4. The ordinate is the $C'/A$ ratio necessary for 75% correct identification from a $C/A$ ratio of 0 db. The lower abscissa is the attenuation, in decibels, imposed upon both the $x$ and $y$ coordinates of the oscilloscope. The upper abscissa presents the relative size of the scatterplot. The parameter is the duration of the display.

Under the conditions of testing, 5- to 10-fold linear reductions in size were easily tolerated, but substantially higher $C'/A$ ratios were required for larger size reductions.

*Series V: Comparison of Positive and Negative Correlations*

A further comparison between auditory and visual tests is furnished by the identification of positive vs. negative correlations. In the auditory tests (Pollack & Trittipoe, 1959), it was observed that only limited use could be made of the direction of the interaural correlation in the vicinity of zero correlation. Specifically, it was found that the absolute value of the correlation necessary to achieve a fixed identification level, $|r_i|$ was nearly identical under the following three testing procedures: Procedure A($+r_i$ vs. $r_o$ or zero correlation); Procedure B ($-r_i$ vs. $r_o$) and Procedure C($+r_i$ vs. $-r_i$).

With visual correlations, nearly complete contribution of the direction of the correlation might be expected. That is, identification of a positive tilt of the correlational scatterplot from a negative tilt should be substantially better than identification of a positive tilt from a zero tilt. Series V tests this point.

In all, four procedures were examined in Series V. Procedures A, B, and C duplicated the corresponding test conditions for the auditory tests.

And, since Procedures A and B yielded equivalent results, only the average of their results will be presented as Procedure A:B. In addition, Procedure D($\pm r_i$ vs. $r_o$) was examined to determine whether uncertainty of the direction of the correlation would retard performance. In Procedure D in a random order, $r_o$ was presented upon one-half of the trials, $+r_i$ was presented upon one-quarter of the trials, and $-r_i$ was presented upon one-quarter of the trials. Only two responses were available to the S: "stubbier" (for $r_o$) and "thinner" (for $r_i$, irrespective of the direction of the correlation).

Each procedure was examined at three durations: 10, 100, and 1000 msec. Table 2 presents the results of Procedures A:B, C, and D of Series V. The first column presents the noise burst duration, in milliseconds. The second column represents interpolated identification levels obtained from appropriate psychometric functions. The third, fourth, and sixth columns present the $r_i^2$ required for the stated level of identification for Procedures A:B, C, and D, respectively. (The initial decimal point has been omitted.) For example, Table 2 says that for 100-msec. noise bursts in Procedure C, $+r^2$ of .045 when paired with $-r^2$ of .045 yielded 75% correct identifications. The ratio between the corresponding entries for Procedures A:B and C is presented in the fifth column.

Table 2 makes three points: (*a*) Procedure C (positive vs. negative correlation) yields the highest sensitivity in that the required correlational differences are substantially lower than obtained with the other procedures. (*b*) The ratio of the correlations (squared) yielded by Procedures A:B and C in the fifth column of Table 2, is nearly 2 to 1, especially for intermediate levels of identification. This result suggests the correlation (squared) measure is an appropriate measure of the correlational variable in that it is additive about zero correlation. (*c*) Uncertainty of direction of correlation associated with Procedure D has a small, but significant, effect in reducing accuracy of identification.

TABLE 2

SUMMARY OF RESULTS WITH POSITIVE AND NEGATIVE CORRELATIONS

| Duration Msec. | % Corr. | Procedure | | | |
|---|---|---|---|---|---|
| | | A:B | C | $\frac{A:B}{C}$ | D |
| 10 | 95 | 376 | 237 | 1.6 | 410 |
| | 90 | 295 | 184 | 1.6 | 343 |
| | 85 | 245 | 143 | 1.7 | 285 |
| | 80 | 205 | 112 | 1.8 | 237 |
| | 75 | 174 | 089 | 2.0 | 195 |
| | 70 | 143 | 072 | 2.0 | 161 |
| | 65 | 119 | 056 | 2.1 | 122 |
| | 60 | 097 | 025 | 3.9 | 089 |
| 100 | 95 | 195 | 137 | 1.4 | 192 |
| | 90 | 155 | 087 | 1.8 | 152 |
| | 85 | 127 | 065 | 2.0 | 130 |
| | 80 | 104 | 054 | 1.9 | 112 |
| | 75 | 084 | 045 | 1.9 | 097 |
| | 70 | 067 | 035 | 1.9 | 081 |
| | 65 | 046 | 026 | 1.8 | 063 |
| | 60 | 029 | 017 | 1.7 | 043 |
| 1000 | 95 | 164 | 074 | 2.2 | 149 |
| | 90 | 124 | 049 | 2.5 | 125 |
| | 85 | 099 | 038 | 2.6 | 097 |
| | 80 | 075 | 031 | 2.4 | 082 |
| | 75 | 056 | 026 | 2.2 | 069 |
| | 70 | 040 | 022 | 1.8 | 056 |
| | 65 | 024 | 018 | 1.3 | 045 |
| | 60 | 013 | 013 | 1.0 | 033 |

### *Series VI: Effect of Sampling Pulse Frequency*

Series I–V considered a correlational display in which a continuously moving electron stream defined the distribution. Series VI–VIII considered the equivalent noise distributions but with the trace visible only in terms of discrete dots spaced in

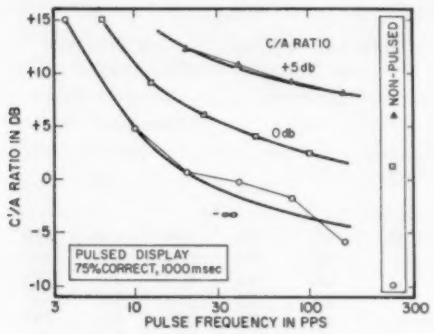



FIG. 5. Effect of discrete pulsing frequency upon identification of two correlations. The abscissa is the pulsing frequency, in pulses/second. The ordinate is the $C'/A$ ratio in decibels required for 75% correct identification from each of three reference $C/A$ ratios (listed as the parameter). For comparison, the results of Table 1 for the nonpulsed display of 1000 msec. are presented. The duration of the display was 1000 msec. Tests of Series VI.

time. This was accomplished by briefly pulsing the $z$ axis of the oscilloscope. The number of samples from the distribution is equal to the number of pulses delivered to the $z$ axis within the available display interval. The number of samples is the product of the pulsing rate (impulses per second) and the available display duration (in seconds). In terms of a statistical hypothesis tester, $S$'s performance would be expected to improve as the size of the sample is increased.

Figure 5 presents the effect of varying the pulsing rate with a display of available duration constant at 1.0 sec. For comparison, results under corresponding nonpulsed conditions are also presented from Fig. 2.

Performance improves substantially with higher pulsing frequencies.

### *Series VII: Effect of Sampling Pulse Number*

The total number of samples, as represented by the total number of pulses per observation interval, is determined by the product of the pulsing frequency and the available duration of the display. Which of the three parameters—number of pulses, pulsing frequency, or available duration of the display—is the critical determinant of performance?

Figure 6 presents the results of tests carried out over a wide range of available display durations and pulsing rates. The abscissa is the available duration of the display; the ordinate is the $C'/A$ ratio in decibels required for 75% correct identification; the parameter is the number of samples (pulses) within the display duration. For purposes of comparison, the corresponding results for the nonpulsed display from Fig. 2 are also presented.

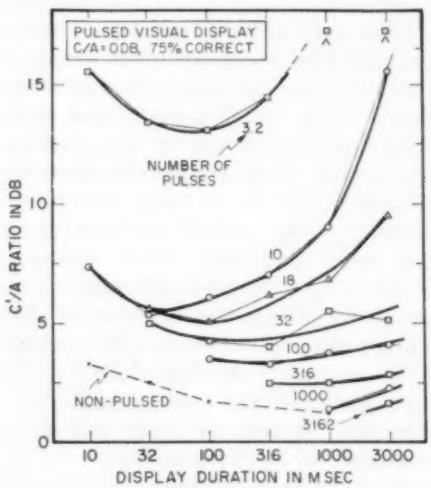



FIG. 6. Effect of number of discrete samples upon identification of two correlations. The abscissa is the display duration in milliseconds. The ordinate is the $C'/A$ ratio required for 75% correct identification from a $C/A$ ratio of 0 db. The parameter is the number of samples produced by pulsing the $z$ axis of the oscilloscope. For comparison, the results of Table 1 for the nonpulsed display are presented as the dashed curve. Tests of Series VII.

Over a wide range of conditions, performance is nearly invariant with a constant number of samples per presentation.[4] The notable exception is the rapid deterioration of discrimination for displays with a small number of pulses. With this exception in mind, we conclude that identification of pulsed correlational displays is, in large part, independent of the display duration and pulsing rate, except insofar as these factors determine the total number of sampling pulses.

### *Series VIII: Effect of Sampling Pulse Duration*

The number of sampling pulses was considered to be the critical determinant of performance in Series VII. However, it may be noted that the effect of sampling pulse number is confounded by the sampling pulse duration. (The total sampling duration is the product of the total number of sampling pulses by the duration per sampling pulse.) In order to test for confounding, additional identification tests were carried out with sampling pulse durations from 2 to 80 μsec with sampling pulse numbers of 3 to 80 pulses per display. It may be noted that, under the longer sampling pulse durations, large segments of the correlated 100–6800 cps wave forms were obtained rather than the discrete dots observed with the shorter pulsing durations.

Over the range of pulse durations examined, equivalent performance was observed for a constant number of sampling pulses, irrespective of the sampling pulse duration. This result lends support to Series VII which concluded that the number of sampling pulses was the critical determinant of performance.

[4] The rapid stabilization of the correlation between 4 and 36 sampling pulses is illustrated by Licklider and Dzendolet (1948, Fig. 2C, p. 122).

## Discussion

The deterioration in the identification of correlational scatterplots with reduction in (*a*) display duration, (*b*) noise bandwidth, and (*c*) pulsing number, all lend support to the interpretation of *S*'s task as that of testing alternative statistical hypotheses.

In this light, the deterioration of performance is not to be attributed to a change in *S*'s performance, but rather to a reduction in the reliability of the statistical information available to the display. Under comparable conditions, the performance of an ideal statistical observer would also suffer. Supporting evidence that statistical sampling of the stimulus was often a controlling factor upon performance is furnished by the extremely high inter-*S* agreement with short nonpulsed samples or with a small number of pulses. Under these circumstances, all *S*s were often in perfect agreement in selecting between the two correlations and often incorrectly.

We have examined several models based on the sampling distribution of the correlation coefficient. To date, we have been singularly unsuccessful in obtaining a model which yields satisfactory quantitative predictions of the data. We strongly suspect that the theoretical development of a performance model for correlational distributions for the ideal statistical observer will point the way to the required model of the human observer.

One final point: it is an interesting exercise to inquire about an appropriate representation of the correlational variable within a discrimination space. Distance in discrimination space is scaled in terms of the confusability among the items: the greater the confusability between two conditions, the closer is their representation in discrimination space. This approach suggests that we might profitably search for an appropriate metric of confusability

such as the $d'$ measure of the theory of signal detectability.

An alternative approach may be employed for two special cases. If the absolute value of correlations required for a given level of identification $|r_i|$ is identical in the testing of ($+r_i$ vs. $r_o$ or zero correlation); ($-r_i$ vs. $r_o$); and ($+r_i$ vs. $-r_i$), we have three points equidistant in discrimination space: $+r_i$, $-r_i$, and $r_o$. The appropriate representation is two discrimination axes, labeled $+r_i$ and $-r_i$ terminating at $r_i$ and separated by an angle of 60°. This result is, in fact, obtained with interaural auditory noise correlations. Direction of correlation may be discriminated, but its effects are only partially additive.

Another special case is observed when the effect of the direction of correlation is completely additive. That is, with respect to an appropriate metric, we need twice the value of the metric required for ($+r_i$ vs. $r_o$) or ($-r_i$ vs. $r_o$) as for ($+r_i$ vs. $-r_i$). In the present examination of the visual correlational scatterplot, $r_i^2$ appears to be the appropriate metric, at least for intermediate identification levels. The representation of the visual scatterplot in discrimination space is, then, simply a single discrimination continuum, scaled in terms of $r^2$, and symmetrical about $r_o$.

## Summary

Visual correlation scattergrams were obtained by mixing a common noise source, $C$, with independent noise sources, $A$ and $B$, and displaying the mixtures ($C + A$) and ($C + B$) across the $x$ and $y$ coordinates of an oscilloscope. The correlational distributions were displayed upon the oscilloscope either (*a*) by a continuous beam of electrons or (*b*) by sampling the beam to obtain discrete samplings of the display. The task of $S$ was to identify whether the reference correlation or a higher correlation was presented. The variables examined were the duration, size, and sampling characteristics of the display and the upper cutoff frequency of the noise distributions.

In terms of the magnitude of the difference between the reference and variable correlations required for 75% correct identification: (*a*) Sensitivity is greatest for the highest reference correlations. (*b*) Sensitivity decreases with narrower bandwidths and shorter display times, with rapid changes below burst durations of 100 msec. (*c*) Sensitivity for the sampled discrete pulsed display is determined by the number of sampling pulses, and is independent of the pulsing rate and duration of the display, except insofar as they determine the number of sampling pulses.

The results are interpreted in terms of the task of $S$ as a tester of alternative statistical hypotheses under conditions of varying reliability of the display information.

## References

Licklider, J. C. R. The influence of interaural phase relations upon the masking of speech by white noise. *J. Acoust. Soc. Amer.*, 1948, **20**, 150–159.

Licklider, J. C. R., & Dzendolet, E. Oscilloscopic scatterplots illustrating various degrees of correlation. *Science*, 1948, **107**, 121–124.

Pollack, I., & Trittipoe, W. Binaural listening and interaural noise correlation, *J. Acoust. Soc. Amer.*, 1959, **31**, 1250–1252. (Abstract)

Zerlin, S. The interaural time disparity threshold as a function of the interaural correlation. *J. Acoust. Soc. Amer.*, 1959, **31**, 127. (Abstract)

(Received June 15, 1959)

*Journal of Experimental Psychology*
Vol. 59, No. 6, 1960

# TASK DIFFICULTY AND DRIVE IN VERBAL LEARNING

R. R. STANDISH AND R. A. CHAMPION

*University of Sydney*

In view of the wide disparity in the results of experiments on the effect of variation in drive strength in verbal learning (e.g., Jensen, 1958) the following study was made in an attempt to test the basic essentials of Spence's theory (Spence, 1958) under the least equivocal conditions. A situation was first provided in which the correct response was clearly dominant, the prediction being that the performance of high-drive *S*s would be superior to that of low-drive *S*s with medium-drive *S*s occupying an intermediate position. It was thought that some previous *E*s in their selection of material to be learned may have merely demanded responses which were relatively easy to acquire without ensuring the actual dominance required by the theory. For example, if a "high association" or easy response is given on 20% of a block of trials then it must have been dominated at least momentarily by some incorrect response on 80% of the block of trials. Thus an increase in drive strength might impair the performance of this "easy" response as much as that of a "difficult" response given on only 10% of the same block of trials. The present interpretation therefore favors clear dominance of the correct response rather than relative ease or high association value if the performance of a high-drive group is to be superior.

A second type of situation was devised to ensure that an incorrect response was clearly dominant at the outset of acquisition, the prediction in this case being that the performance of low-drive *S*s would be superior to that of high-drive *S*s up to the point where the clear dominance of the incorrect response was lost, when there would be a reversion to the first type of situation. As before, medium-drive *S*s were expected to occupy an intermediate position. Judging by the results of some other experiments (e.g., Spence, 1956) in which low-drive *S*s have performed better on low association-value or "difficult" responses, this second type of situation is more readily obtainable. The above two predictions are most simply taken together as involving an interaction between task difficulty and drive level.

## METHOD

*Subjects.*—The *S*s were 63 students taking courses in psychology or medicine at the University of Sydney. Three *S*s obtained a "lie score" of 7 or more on the index of drive level, the Taylor Manifest Anxiety Scale (MAS) (Taylor, 1953), and were rejected. The remaining 60 *S*s were each allocated to one of three drive groups on the basis of their anxiety score on the MAS. The 20 low scorers (2–11) constituted the low-drive group (Mean = 7.5, Mdn. = 7.5, $SD$ = 2.4), the 20 high scorers (20–33) the high-drive group (Mean = 24.8, Mdn. = 24.0, $SD$ = 3.4) and the remaining 20 *S*s made up the medium-drive group (Mean = 15.5, Mdn. = 15.5, $SD$ = 2.2).

*Learning materials.*—A list of 10 easy paired-associates was constructed by using common associates in the Kent-Rosanoff list (Rosanoff, 1938). To reduce intralist confusions no stimulus word was used which had been given as a response at any point in the Kent-Rosanoff study; the response word in each case was the answer most commonly given in free association. To ensure clear dominance of the correct response, performance on this easy list was not investigated until each *S* had given the correct response on two successive trials and no measures of perform-

ance (response speed) were used prior to this point. A list of 10 difficult paired-associates was constructed by taking the stimulus words of the easy list and pairing them with response words not given in free association according to the Kent-Rosanoff list. To ensure clear dominance of the incorrect response all *S*s learned both lists, with the difficult list following the easy list.

*Procedure.*—The items were presented in a Ranschburg exposure apparatus with approximately 30 sec. between trials. The *S* was allowed 4 sec. to anticipate the response with the stimulus word present (in blue), and the response word (in red) was then presented for 4 sec. To reduce intralist associations the items of each list were presented in three orders, repeated as necessary. An electric commutator actuated the exposure apparatus and started an electronic chronoscope simultaneously with the presentation of each stimulus word. The chronoscope was stopped by *E* pressing a key as soon as *S* gave the correct response or else it was reset after 4 sec. in readiness for the next stimulus. The easy list was presented to each *S* first, to the criterion of two successive errorless trials followed by eight further trials. Then the difficult list was presented until the same criterion had been reached. Finally, each *S* answered the MAS items.

## Results

The results obtained with speed of correct response (Reciprocal Latency $\times 10^5$) as a measure of performance are presented in Fig. 1. The scores plotted for performance on the easy list are taken from the first two errorless trials and the following eight trials, all of which are labeled as Trials 1–10. The scores for the difficult list are taken from two stages of learning: (*a*) the first 10 trials (labeled also as Trials 1–10) and (*b*) the first two errorless trials and the following eight trials (labeled as Trials A–J). It should be noted that the mean scores for the first two trials on the difficult list are spuriously depressed through the use of an arbitrary maximum response time of 4 sec.

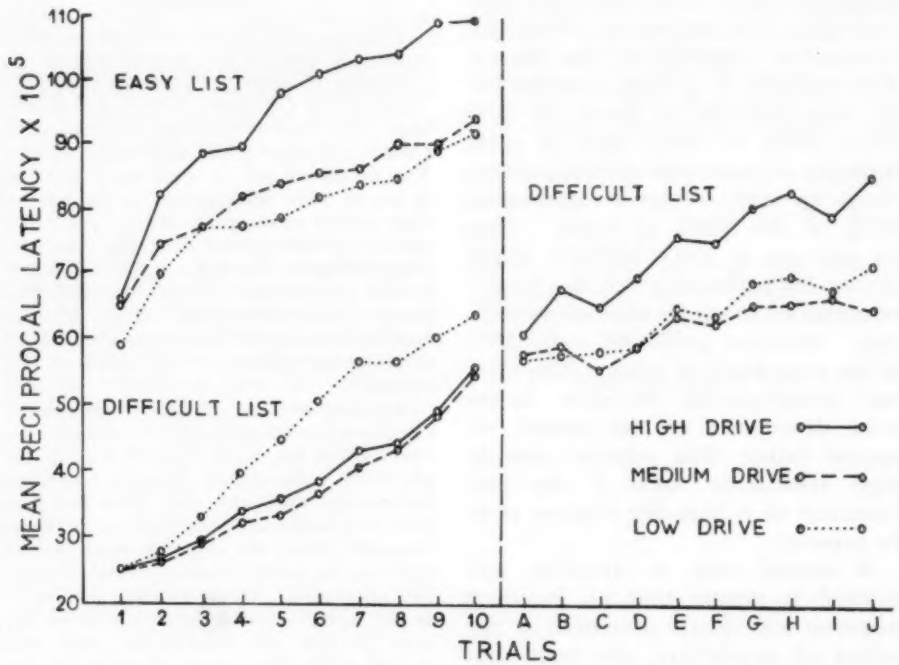



FIG. 1. Mean response-speed early (Trials 1–10) and late (Trials A-J) in learning.

Trials 1–10 as defined above provide the primary data of the experiment. These data are in accord with the prediction for the easy list, with the high-drive group superior in performance to the low-drive group throughout and the medium-drive group falling between these extremes. On the difficult list the prediction is confirmed to the extent that the low-drive group shows superior performance to the high-drive group on all trials except the first, but there is a deviation from prediction in that the high-drive *Ss* perform at a consistently higher level than the medium-drive *Ss* when the opposite result was expected. The results of an analysis of variance in these data are summarized in Table 1, homogeneity of variance having been established. The predicted interaction is statistically significant.

A further test of the general theory under consideration is available in the data for the difficult list over Trials A-J as defined above and in a comparison of these data with the scores from Trials 1–10 for the same list. When correct responses are given on the difficult list it takes on the character of the easy list, leading to the prediction that high-drive *Ss* should now perform at a superior level despite their previous inferiority. That this effect does occur is shown in Fig. 1 although now the low-drive group is slightly superior to the medium-drive group at most stages whereas it would have been expected to be inferior in performance. The effect now under consideration in the comparison of performance on the difficult list at two stages of learning involves an interaction similar to that already considered in Table 1. This was tested in a second analysis of variance, summarized in Table 2. The interaction is statistically significant and the second analysis confirms the finding in the primary data.

A final test of the theory was made with two other measures of performance available in the acquisition of the difficult list, percentage fre-

TABLE 1

ANALYSIS OF VARIANCE IN SPEED OF RESPONSE SCORES OVER TRIALS 1–10

| Source | *df* | *MS* | *F* |
|---|---|---|---|
| Task difficulty (A) | 1 | 588,699.18 | 488.38** |
| Drive level (B) | 2 | 4,148.01 | 3.44* |
| A × B | 2 | 15,655.73 | 12.99** |
| *Ss* | 114 | 1,205.42 | 6.90** |
| Measures | 960 | 175.69 | |

* $P < .05$.
** $P < .001$.

TABLE 2

ANALYSIS OF VARIANCE IN SPEED OF RESPONSE SCORES ON THE DIFFICULT LIST IN EARLY AND LATE STAGES OF LEARNING

| Source | *df* | *MS* | *F* |
|---|---|---|---|
| Stage of learning (A) | 1 | 183,300.84 | 228.85** |
| Drive level (B) | 2 | 4,838.81 | 6.04* |
| A × B | 2 | 10,117.41 | 12.63** |
| *Ss* | 114 | 800.95 | 5.54** |
| Measures | 960 | 144.51 | |

* $P < .01$.
** $P < .001$.

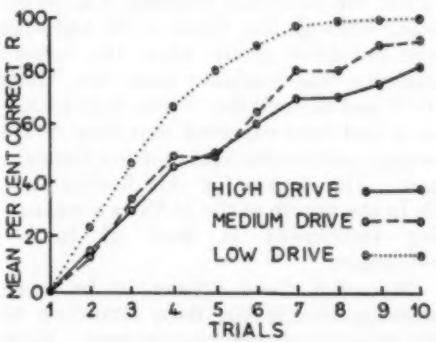



FIG. 2. Mean percentage frequency of correct response for the difficult list early in learning.

quency of correct response and number of trials to the criterion. The result using the first of these measures over Trials 1–10 is plotted in Fig. 2, wherein it may be observed that the relative disposition of the low-, medium-, and high-drive groups deviates only slightly from that predicted. The mean trials-to-criterion scores for the low-, medium-, and high-drive groups were 7.86, 12.25 and 14.85 trials, respectively, and exactly in accord with the prediction.

## Discussion

The results of this experiment are clearly compatible with the theory under consideration, presumably because it was conducted as nearly as possible within the specified boundary conditions. It would therefore seem that the effect of other methods of varying task difficulty in verbal learning, e.g., through intralist similarity, should not necessarily be accepted as support for or disproof of the theory in the present state of knowledge but that an attempt should rather be made to extend the theory to account for more complex circumstances. Some features of the present study suggest factors which might be taken into account in this endeavor.

First, the equivocal results of the medium-drive group deserve attention. These *S*s fell with the high-drive group when the incorrect response was dominant (difficult list, Trials 1–10) and with the low-drive group when the correct response was dominant (easy list, Trials 1–10 and difficult list, Trials A-J) whereas it had been expected that they would occupy an intermediate position throughout. The reason for this finding may lie in the nature of the MAS as a measuring instrument in need of further refinement.

Secondly, there appear to be some discrepancies in the data according to the measure of performance used. With speed of correct response as an index of response strength the performance curves for the acquisition of the difficult list cross at an intermediate value of about 60 units on the scale used (by interpolation in Fig. 1). With percentage frequency of correct response as a measure, however, no such crossing was observed (Fig. 2). An additional phenomenon is the drop in the response speed of the low-drive group on the difficult list at about the stage of the first four errorless trials (Trials A-D, Fig. 1) whereas the other two groups continue to improve; this finding was not apparent with the other measure of performance. These discrepancies are almost certainly due to the fact that response speed is a more sensitive measure and one which does not present an arbitrary maximum level, but it is possible that percentage frequency of correct response involves some complication as yet not made explicit.

Finally, whereas performance on the easy task at Trial 10 when averaged over the three drive groups is 98.3 units on the scale of response speed, that for the difficult task at Trial J is only 73.7 units on the same scale (Fig. 1). The similar disposition of the high- and low-drive groups at these points, as well as the use of the same criterion of performance, would suggest that the absolute level of performance should be the same in the two cases and the obtained difference was not anticipated. A reason for expecting a difference may be found after the event in the possible inhibiting effects of nonreinforcement (Spence, 1956) which might carry over from Trials 1–10 to Trials A-J in the case of the difficult list but which could not be generated in Trials 1–10 with the easy list.

## Summary

Situations were sought in which the basic implications of Spence's theory about the interaction of task difficulty and drive in verbal learning could be tested in the least equivocal fashion. An easy list in which the correct response clearly dominated was obtained by measuring speed of response over a series of errorless trals and a difficult list involving clear dominance of the incorrect

response was achieved by having *Ss* then learn new responses to the stimuli of the easy list. With score on the Taylor MAS as an index of drive level, high- and low- drive groups performed as expected, with the high-drive group superior on the easy list and inferior on the difficult list but medium-drive *Ss* failed to occupy the predicted intermediate position with any consistency. The nature of the MAS as a measuring instrument, the index of response-strength used, and inhibition due to nonreinforcement were indicated as three specific factors contributing to disparity in the results of experiments in this area as well as general failure to keep within the boundary conditions of the type of theory being tested.

## References

JENSEN, A. R. Personality. *Ann. Rev Psychol.*, 1958, **9**, 295–322.

ROSANOFF, A. J. *Manual of psychiatry.* New York: Wiley, 1938.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

SPENCE, K. W. A theory of emotionally based drive (*D*) and its relation to performance in simple learning situations. *Amer. Psychologist*, 1958, **13**, 131–141.

TAYLOR, J. A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

(Received June 19, 1959)

*Journal of Experimental Psychology*
Vol. 59, No. 6, 1960

# ASSOCIATION VALUE AND FAMILIARITY IN SERIAL VERBAL LEARNING[1]

RICHARD H. LINDLEY

*Carnegie Institute of Technology*

A well known generalization about verbal learning is that serial lists of high association value (HAV) nonsense syllables are easier to learn than lists of low association value (LAV). This generalization holds true under fairly varied conditions; e.g., when the methods of complete presentation and unaided recall are used (McGeoch, 1930) or when the syllables are spelled and presented serially on a memory drum (e.g., Underwood & Richardson, 1956).

Analysis of lists of nonsense syllables in which association values were determined (Glaze, 1928; Hull, 1933; Krueger, 1934) shows that some of the syllables are familiar since they are the first three letters of English words (e.g., BEL-BELL; FAS-FAST). Ninety-seven per cent of the 100% association value syllables in Glaze's lists are the first three letters of one or more of the 20,000 most common words in the English language as determined by Thorndike and Lorge (1944). The comparable figures for the 80% and 0% association value syllables are 70% and 0%, respectively. Thus differences in the rate of learning HAV and LAV syllables may simply be due to differences in the familiarity of these kinds of items.

The present study compared the rates of learning of lists of nonsense syllables which had the same published association values but which differed appreciably in familiarity. If the relationship between association value and rate of learning is due to familiarity, the familiar syllables should be easier to learn than the unfamiliar ones regardless of association value.

A second implication of the familiarity hypothesis is that nonsense syllables should be no more difficult to learn than meaningful words provided that both kinds of items are equally familiar. McGeoch (1930) found that words were easier to learn than HAV syllables; however, no control over familiarity was reported in that study. The present study controlled for familiarity.

## METHOD

*Experimental design.* — There were four main groups of *Ss*. Group I learned a list of meaningful words; Group II learned familiar HAV syllables; Group III learned unfamiliar HAV syllables; and Group IV learned unfamiliar LAV syllables. The study consisted of two replications; the first used syllables drawn from Glaze's lists and the second used syllables drawn from Krueger's lists. Twelve *Ss* were assigned to each group in each replication, making a total of 24 *Ss* in each main group.

In the first replication the meaningful words were constructed by adding letters to the familiar HAV syllables to form words (e.g., FOL-FOLD). The results showed almost no difference in the rate of learning of the familiar HAV syllables and words. The failure to obtain a significant difference could have been due to the fact that the words were longer than the syllables; therefore, *Ss* in Group I in the first replication were replaced with *Ss* who learned words of the same length

[1] This paper was presented at the Midwestern Psychological Association Convention in 1959. The research was supported in part by a grant from the Humanistic and Social Studies Division of Carnegie Institute of Technology. Thanks are due Mary Ann Turnquist for clerical assistance.

TABLE 1

LISTS USED

| Words (Repl. 1 and 2) | Glaze's Syllables (Repl. 1) | | | Krueger's Syllables (Repl. 2) | | |
|---|---|---|---|---|---|---|
| Group 1[a] | Group 2[a] | Group 3 | Group 4 | Group 2[a] | Group 3 | Group 4 |
| LIP | WID | RUJ | LEQ | WES | LUV | YUT |
| TAX | FOL | WIM | CIJ | BUR | QIT | BIW |
| DOG | JAC | BOC | WUB | PIC | ROX | GOX |
| RED | KEP | DEY | ZOM | GOV | WUN | DAJ |
| CUT | BIR | LAQ | DAX | NAM | DAC | ZIQ |
| MAN | HOM | VIF | NEH | JUL | YEZ | FAP |
| FEW | SUG | TAZ | VAF | DIF | KOS | KEC |
| JOY | NEX | NOK | YOP | YEL | JEM | VUH |
| HIS | GAV | HUP | GUK | TAK | FAX | MOJ |

[a] All items for Groups 1 and 2 occurred 100 times per million or more in the Thorndike-Lorge count except TAX, JAC, and JUL which occurred 50 times per million but less than 100 times.

as the nonsense syllables. The latter *Ss* were run during the second replication.

*Materials.*—Familiarity was estimated by using the G or General count of the Thorndike and Lorge (1944) word count. The familiar syllables were the first three letters of English words which occurred 50 times per million or more in English text. The unfamiliar syllables were not the first three letters of any words occurring one or more times per million. Thirteen of these syllables were not parts of any words which occurred 10 times per million or more; five of the syllables occurred once or twice in positions other than the first three letters in words which occurred 10 times per million or more. Thus the familiar HAV syllables were clearly more familiar than the unfamiliar HAV syllables. The meaningful words were selected from three-letter words in the Thorndike-Lorge count which occurred 50 times per million or more.

Each list contained nine items. The items were selected with the restriction that formal intralist similarity be low. No consonant letter appeared more than twice in any one list. The items are presented in Table 1 in the orders in which they were learned by *Ss*. The mean association values and ranges of Glaze's syllables for Group II were 84.6% and 67%–100% respectively; for Group III, 84.6% and 67%–100%; for Group IV, 7.4% and 0%–27%. The corresponding values of Krueger's syllables for Group II were 94.5% and 85.5%–99%; for Group III, 94.4% and 89.5%–96%; for Group IV, 49.4% and 34%–60%.

*Procedure.*—The items were presented on a memory drum at a 2-sec. rate. An asterisk served as a cue for the first item. There was a 6-sec. pause between trials. Typical instructions for learning by the anticipation method when the items are spelled were read to *S* at the start of the experiment. The lists were learned for 25 trials and the mean number of correct responses occurring during these trials constitute the major data of this study.

At the termination of serial learning, *S* was instructed to report any meaningful associations he had made to the nonsense syllables during serial learning. The *E* spelled the syllables one at a time and *S* reported any associations he had made.

*Subjects.*—The *Ss* were 96 undergraduate men and women from Carnegie Institute of Technology. None of the *Ss* had had previous nonsense syllable learning experience. The *Ss* were assigned to groups according to a randomized block design.

## RESULTS

*Serial learning.*—The means and *SDs* for the mean number of correct responses during the 25 serial learning trials are presented in Table 2. Analysis of variance showed that only the differences between groups were significant ($F = 19.90$, $df = 3$ and 66, $P < .01$). Since the $F$ for replications was not significant, the means for the individual groups were pooled and the differences between the main groups were tested

TABLE 2

MEANS AND *SD*s FOR NUMBER OF CORRECT RESPONSES IN 25 TRIALS

| Groups | Replication 1 | | Replication 2 | |
|---|---|---|---|---|
| | Mean | *SD* | Mean | *SD* |
| I | 166.25 | 15.69 | 169.25 | 22.89 |
| II | 153.67 | 45.43 | 151.83 | 34.48 |
| III | 116.17 | 37.80 | 123.25 | 51.80 |
| IV | 103.67 | 30.95 | 83.00 | 27.18 |

TABLE 3

MEAN NUMBER CORRECT PER TRIAL FOR SUCCESSIVE BLOCKS OF FIVE TRIALS FOR THE MAIN GROUPS

| Groups | Trial Blocks | | | | |
|---|---|---|---|---|---|
| | 1–5 | 6–10 | 11–15 | 16–20 | 21–25 |
| I | 3.49 | 5.93 | 7.33 | 8.30 | 8.34 |
| II | 3.26 | 5.45 | 6.61 | 7.28 | 7.77 |
| III | 2.32 | 3.99 | 4.80 | 6.18 | 6.69 |
| IV | 1.44 | 2.63 | 3.80 | 4.98 | 5.85 |

with the $t$ test; a difference of 22.01 is significant at the 5% level. Thus all differences reached $P = .05$ except that between Groups I and II.

Inspection of Table 2 shows that there is a trend for the meaningful words to be easier to learn than the familiar HAV syllables. Further analysis of the data showed that four *S*s in Group II failed to learn the list to a criterion of one perfect recitation within 25 trials whereas all *S*s in Group I learned the list within 25 trials. When medians are compared, the difference between Groups I and II disappears. (The median number correct was 166 for Group I, 163 for Group II, 124.5 for Group III, and 93.5 for Group IV.) In terms of the number of trials to reach a criterion of one perfect recitation, the mean number of trials was 11.83 for Group I ($N = 24$) and 11.65 for Group II ($N = 20$). Thus, for the majority of *S*s, the familiar HAV syllables were no more difficult to learn than the meaningful words.

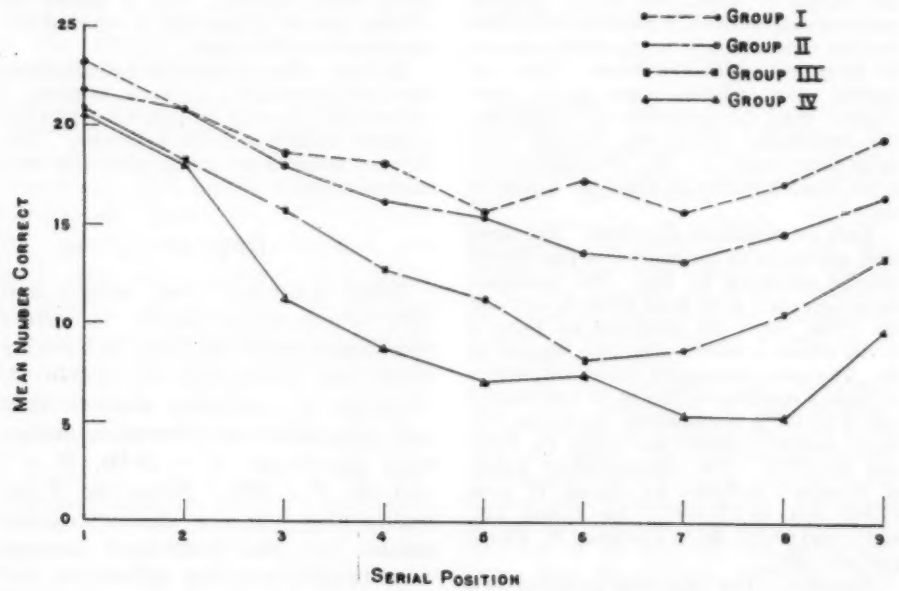



FIG. 1. Mean number correct in 25 trials as a function of serial position.

TABLE 4

MEANS AND *SD*s FOR NUMBER OF ASSOCIATIONS

| Groups | Replication 1 | | Replication 2 | |
|---|---|---|---|---|
| | Mean | *SD* | Mean | *SD* |
| II | 6.33 | 2.49 | 7.00 | 1.78 |
| III | 4.33 | 2.17 | 5.67 | 1.65 |
| IV | 4.08 | 2.29 | 4.00 | 2.16 |

The mean number of correct responses per trial for successive blocks of five trials for the main groups are presented in Table 3. The curves for the high familiarity items are negatively accelerated and asymptotic to the $x$ axis while the curves for the low familiarity items approximate straight lines. This may simply reflect the fact that the high familiarity items were better learned during the 25 trials than the low familiarity items. Figure 1 presents serial position curves for the four main groups. Skewed, bowed serial position curves which are typical of learning disconnected verbal material were obtained for all groups. The curves for the high familiarity items are less bowed than the curves for the low familiarity items.

*Number of associations.*—The means and *SD*s for the number of associations reported by *S*s are presented in Table 4. Analysis of variance showed that only the differences between groups were significant ($F = 9.99$, $df = 2$ and 44, $P = .01$). The $t$ test shows that a difference of 1.23 associations between groups will be significant at $P = .05$. Thus only the differences between Groups II and III, and II and IV reach an acceptable level of significance.

## DISCUSSION

The results clearly show that familiarity is an important variable in serial nonsense syllable learning. Even though the published association values of the syllables learned by Groups II and III were equal, the familiar syllables were significantly easier to learn than the unfamiliar syllables. Furthermore, when familiarity was controlled, the familiar syllables were no more difficult to learn than the meaningful words for most *S*s.

The results also showed that there was a direct relationship between ease of serial learning and the number of associations evoked by the syllables, although only the differences in associations between the familiar and unfamiliar syllables reached the .05 level of significance. These results are consistent with the theories of Mandler (1954), and Dollard and Miller (1950) which predict that items which elicit distinctive implicit responses will be easier to learn than items which do not elicit such distinctive responses. The *S*s in the present study reported more associations (distinctive responses) to the familiar syllables than they did to the unfamiliar syllables. The fact that the unfamiliar HAV syllables were easier to learn than the LAV syllables can be explained by this reasoning. That is, the unfamiliar HAV syllables elicited more associations than the LAV syllables and thus may have been more distinctive.

The results of this study are consistent with results obtained by Hall (1954). In that study familiar words were found to be easier to recall than unfamiliar ones. Thus familiarity is an important variable in the learning of both nonsense syllables and words.

## SUMMARY

The present study tested the hypothesis that the relationship obtained between published association values and ease of serial learning was due to familiarity. The results showed that lists of familiar syllables were easier to learn than lists of unfamiliar syllables when the items were spelled, even though the average published association values for the lists were identical. Furthermore, no significant difference was obtained between the rate of learning of meaningful words and nonsense syllables when familiarity was equated. The average number of associa-

tions evoked by the syllables during learning was directly related to ease of learning. The results were interpreted as being consistent with theories which emphasize the role of distinctive implicit responses in verbal learning.

## REFERENCES

DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.

GLAZE, J. A. The association value of nonsense syllables. *J. Genet. Psychol.*, 1928, **35**, 255–269.

HALL, J. F. Learning as a function of word-frequency. *Amer. J. Psychol.*, 1954, **67**, 138–140.

HULL, C. L. The meaningfulness of 320 selected nonsense syllables. *Amer. J. Psychol.*, 1933, **45**, 730–734.

KRUEGER, W. C. F. The relative difficulty of nonsense syllables. *J. exp. Psychol.*, 1934, **17**, 145–153.

MANDLER, G. Response factors in human learning. *Psychol. Rev.*, 1954, **61**, 235–244.

MCGEOCH, J. A. The influence of associative value upon the difficulty of nonsense-syllable lists. *J. Genet. Psychol.*, 1930, **37**, 421–426.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words.* New York: Bureau Publ., Columbia Univer., 1944.

UNDERWOOD, B. J., & RICHARDSON, J. A. The influence of meaningfulness, intralist similarity, and serial position on retention. *J. exp Psychol.*, 1956, **52**, 119–126.

(Received June 23, 1959)

*Journal of Experimental Psychology*
Vol. 59, No. 6, 1960

# EFFECTS OF PROBABILITY OF MISINFORMATION AND NUMBER OF IRRELEVANT DIMENSIONS UPON CONCEPT IDENTIFICATION [1]

VLADIMIR PISHKIN[2]

*University of Utah*

The primary aim of the present experiment was to investigate the effect of incorrect, or misinformative, feedback (MF), in conjunction with various degrees of task complexity, upon the performance of human *Ss* in concept identification.

Variation of misinformative feedback, MF, in this experiment, was an operation similar to partial reinforcement in a two-choice guessing problem, where asymptotic performance typically reaches probability matching. The present experiment differed from guessing game studies (e.g., Grant, Hake, & Hornseth, 1951; Humphreys, 1939) in the following respects: (*a*) percentage of correct feedback for either of the two responses depended upon only one binary dimension of the stimulus, and (*b*) stimulus information irrelevant to either response was introduced in a quantifiable manner. It was hypothesized that the introduction of irrelevant information would reduce rate of learning. Despite differences in methodology, it was expected that probability matching would take place in this experiment as in other partial reinforcement studies. Morin (1955) has shown that the acquisition of correct responses in a two-choice discrimination situation is negatively and progressively influenced by MF. Morin's data suggest that various degrees of MF in a concept identification situation would produce differential acquisition rates of correct responses.

Estes and Straughan's (1954) model demonstrated that with the exception of a few deviant cases, terminal response probability approached theoretical asymptotic response probability, $\pi$, the probability of reinforcement for that response. Restle (1955, 1957) has proposed a model similar to Estes and Straughan's except that the proportion of relevant stimulus cues in a problem determines the rate of learning. This theory assumes a population of available stimuli consisting of sets of cues which are either relevant or irrelevant to the problem's solution. Formally, $\delta = r/r + i$, where $r$ is number of relevant cues, $i$ is number of irrelevant cues, and $\delta$ is the learning rate. By assuming that over a series of trials all irrelevant cues will be adapted and thus rendered ineffective, and that probability of response equals proportion of conditioned, unadapted cues, Restle was somewhat successful in predicting the outcome of a variety of two-choice discrimination problems. In Restle's (1955) framework, the number of errors ($E_t$) made before solution is

[1] This article is based on a dissertation submitted to the Department of Psychology of the Graduate School of the University of Utah in partial fulfillment of the requirements for the PhD degree. The author is especially indebted to Lyle E. Bourne, Jr., for his guidance and assistance throughout the course of this investigation. This research was supported, in part, by a research grant (M-1759) from the National Institute of Mental Health, United States Public Health Service.

[2] Now at Research Laboratory, Veterans Administration Hospital, Tomah, Wisconsin.

approximated by:

$$E_t \cong \tfrac{1}{2} + \tfrac{1}{2}\left[\frac{\log \theta}{(1-\theta)\log(1-\theta)}\right] \quad [1]$$

where $\theta$ is a rate of learning parameter, originally assumed equal to $\delta$. In a later paper, Restle (1957) reasoned that rate of learning parameters depend on $\pi$, the proportion of time a relevant cue is reinforced. Further, Restle proposed that a relevant cue which is reinforced less than 100% of the time has validity less than unity. The following equation represents cue validity, $V$, as a function of $\pi$:

$$V = 4\pi^2 - 4\pi + 1 \quad [2]$$

Then, in the extended model:

$$\theta = V\delta \quad [3]$$

Since $S$ may not perceive all of the cues presented, the parameter $\delta$ must be estimated from data. Once this has been done, $E_t$ should be predictable for various conditions in the present design. Thus, this study serves, in part, as a test of Restle's model.

The threefold purpose of the present experiment was: (*a*) to determine the effect of MF under various degrees of complexity, (*b*) to test the hypothesis of probability matching, and (*c*) to account for the data in terms of Restle's discrimination model.

## Method

*Subjects.*—The *S*s were 180 volunteer students in elementary psychology courses who were divided randomly into 45 treatment groups of 4 *S*s each (2 men and 2 women).

*Task and apparatus.*—All *S*s were presented a series of 192 geometric patterns with a Dunning animatic projector. The screen, mounted at *S*'s eye level, was bordered by an opaque shield which prevented *S* from seeing other than pertinent apparatus. In front of *S* was a panel with two response keys and two amber signal lamps. The *S*'s task was to solve a two-choice discrimination problem by successively matching one or the other of two keys to invariant pattern dimensions, e.g., depression of the left key for the image of a triangle; of the right, for a square, etc. The keys were not marked in any way.

*Procedure.*—At the beginning of each session, *S* was verbally instructed as to the nature of the task, the meaning of the feedback lamps, number of trials required, and manipulation of the controls. The *S* began his task by viewing a geometric pattern projected on the screen. After a self-determined interval, *S* responded by pressing one of the two keys. The momentarily depressed key initiated a signal which recorded *S*'s choice of responses on the chart of an Esterline-Angus operations recorder and triggered a cascaded multiple-phase electronic timer.

Phase 1 of the timer advanced the filmstrip to a blank frame. Phase 2 activated a Western Union tape transmitter precoded with a tape which matched the filmstrip programming, thereby enabling the information feedback, irrespective of *S*'s choice, to be registered on the Esterline-Angus operations recorder. In addition, the same information coded on the tape led to the illumination of one of the two panel lamps above *S*'s keys for a duration of 1 sec. Phase 3 advanced the filmstrip to the next stimulus permitting *S* to begin another trial 5 sec. after *S*'s last response.

Nine strip-filmed series of patterns were used. There were three different problem types involving form, size, or number, as the relevant dimension. A dimension was termed "relevant" when its use by *S* was requisite to identify a pattern with an appropriate key. Each problem type was further elaborated into one of three specific problems which differed in complexity, i.e., amount of irrelevant information. All irrelevant dimensions had a zero correlation with correct responses. When a dimension was neither relevant nor irrelevant, it appeared without variation at only one of its two levels within a given series.

In addition to the basic variations of relevant and irrelevant dimensions, some *S*s performed under a third condition wherein they received erroneous feedback. This feedback is interpreted by the *E* as MF. The probabilities that *S* would receive a "correct" signal over the 192 trials when he made a response actually correct with respect to the relevant dimension, were 1.00, .90, .80, .70, and .60 as determined by 0, 10, 20, 30, and 40% of MF, respectively. That is, on some occasions when *S* had made a "correct"

TABLE 1

ANALYSIS OF VARIANCE ON ERRORS

| Source | *df* | *MS* | *F* |
|---|---|---|---|
| Misinformation feedback (MF) | 3 | 452.58 | 50.45*** |
| Linear | 1 | 1,357.23 | 151.31*** |
| Quadratic | 1 | .13 | |
| Cubic | 1 | .40 | |
| Complexity (C) | 2 | 38.97 | 4.34* |
| Linear | 1 | 76.26 | 8.50*** |
| Quadratic | 1 | 1.67 | |
| Problems (P) | 2 | 6.15 | |
| Sex (S) | 1 | .55 | |
| MF × C | 6 | 56.58 | 6.31*** |
| Linear | (3) | 64.27 | 7.16*** |
| Linear (1) | (1) | 42.60 | 4.75* |
| Linear (2) | (1) | 133.01 | 14.83*** |
| Linear (3) | (1) | 17.25 | 1.92 |
| Quadratic | (3) | 48.88 | 5.38** |
| Quadratic (1) | (1) | 4.55 | |
| Quadratic (2) | (1) | 140.28 | 15.64*** |
| Quadratic (3) | (1) | .01 | |
| MF × P | 6 | 6.66 | |
| MF × S | 3 | 7.68 | |
| C × P | 4 | 5.96 | |
| C × S | 2 | 11.57 | 1.29 |
| P × S | 2 | 4.32 | |
| MF × C × P | 12 | 16.01 | 1.78 |
| MF × C × S | 6 | 12.95 | 1.44 |
| MF × P × S | 6 | 12.67 | 1.41 |
| C × P × S | 4 | 16.09 | 1.79 |
| MF × C × P × S | 12 | 18.48 | 2.06* |
| Within | 72 | 8.97 | |

* $P < .05$.
** $P < .01$.
*** $P < .001$.

response with respect to the relevant dimension, the signal lamp indicating a "wrong" response was lighted, deliberately misinforming *S* as to his choice of keys. Conversely, on some occasions, *S* received feedback to indicate he had responded correctly when, in fact, he had not done so. The scheduling of MF was determined by use of a table of random numbers. The order of pattern presentations was semirandom (i.e., all patterns appearing equally often but with the restriction that the same pattern was never presented on successive trials).

*Design.*—A 5 × 3 × 3 × 2 factorial design was used which incorporated the five degrees of MF (0, 10, 20, 30, or 40% of *S*'s responses met with erroneous feedback over 192 trials), 3 levels of complexity (1, 3, or 5 irrelevant dimensions), 3 problems (form, square, or triangle; size, large or small; or number, one object or two objects, as the relevant dimension), and 2 sex groups.

## RESULTS

The dependent variable studied was number of errors with respect to the *E*-defined relevant dimension over the 192 stimulus patterns. This score was analogous to the dependent variable used in guessing games and is not the number of errors as seen from *S*'s point of view.

*Main effects.*—The data were analyzed for the following four main effects: MF, Complexity, Problems, and Sex. The 192 stimuli were pooled into 12 blocks of 16 patterns each for the convenience of statistical and graphic treatment. Owing to the visible character of the graphs and the differences in learning rate, the analysis of variance was performed on the last two blocks (32 trials), omitting the Blocks variable and the 0% MF group. Summary of this analysis is presented in Table 1. Only the main effects of MF and Complexity were found to be significant. The increase of mean number of errors over the five degrees of MF, which was apparently linear, is shown in Fig. 1. Orthogonal polynomial analysis was performed on this function and only the linear component reached a significant level ($P < .001$). The significance of Complexity agrees with previous work on effects of this

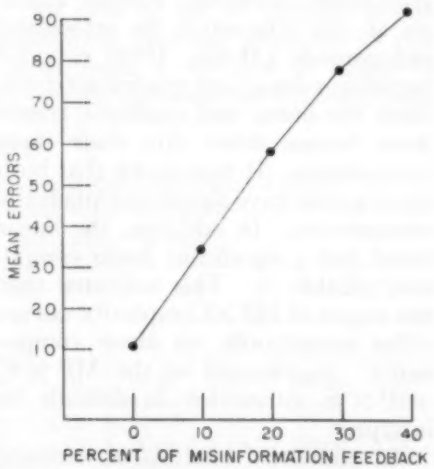



FIG. 1. Mean errors as a function of percentage of MF.

variable in concept identification (Archer, Bourne, & Brown, 1955; Bourne, 1957; Bourne & Pendleton, 1958; Brown & Archer, 1956). The mean numbers of errors for the conditions of 1, 3, and 5 irrelevant dimensions for the two blocks were 7.79, 8.73, and 10.31, respectively. The orthogonal polynomial analysis verified that only the linear term of Complexity reached significance ($P < .001$).

As expected, the main effect of Sex was insignificant. Neither was the Problems main effect significant, a result inconsistent with those of Archer, Bourne, and Brown (1955), and Brown and Archer (1956). The fact that the same irrelevant dimensions were used with the three relevant dimensions, supports the results of Grant and Curran (1952), and Heidbreder (1948, 1949), where identification of number and form was found to be equally difficult.

Following the suggestion of Lindquist (1953), graphic presentations of interactions were made. The interaction of MF × Complexity indicated that MF has a greater retarding influence on performance as task complexity increases. Further analysis of this interaction by orthogonal polynomials (Myers, 1959) revealed significant linear and quadratic trends. Both the linear and quadratic trends were broken down into their three components. It was shown that both main trends have significant quadratic components. In addition, the linear trend had a significant linear component (Table 1). This indicated that the slopes of MF × Complexity curves differ significantly on these components. Significance of the MF × C × P × S interaction is difficult to interpret.

An additional analysis of variance was performed on Blocks. This revealed a significant inverse relationship between number of trials and errors ($F = 41.62$, 11 and 990 $df$, $P < .01$). The interaction of Blocks × MF was also significant ($F = 4.29$, 44 and 990 $df$, $P < .01$), indicating that *S*s performing at the varying levels of MF approach different terminal levels of performance.

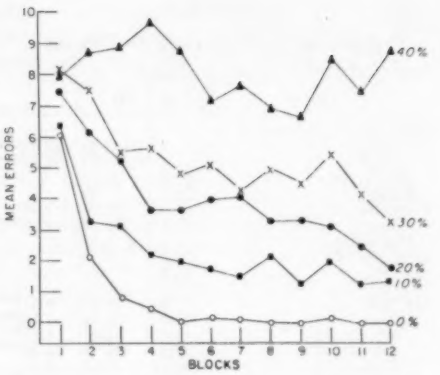



FIG. 2. Mean errors per block of 16 patterns. The parameter is percentage of MF. One irrelevant dimension.

*Probability matching.*—Figures 2, 3, and 4 represent mean number of errors per block for each MF group at each level of complexity. The obtained asymptotic value for each of the four degrees of MF was based on the mean number of errors on

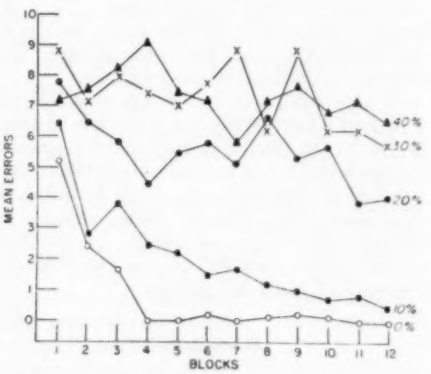



FIG. 3. Mean errors per block of 16 patterns. The parameter is percentage of MF. Three irrelevant dimensions.

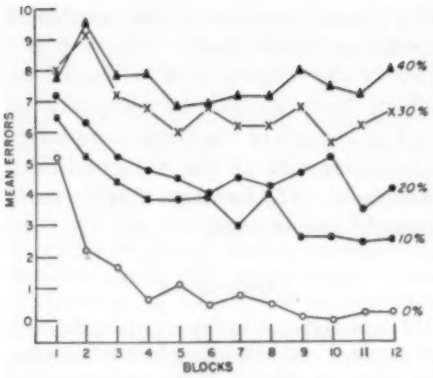



FIG. 4. Mean errors per block of 16 patterns. The parameter is percentage of MF. Five irrelevant dimensions.

Blocks 11 and 12, subgroups being pooled over the insignificant Sex and Problems variables. The expected error scores (probability matching) were determined by the probability of MF. Since there were 32 trials in the last two blocks, the expected error scores were 3.2, 6.4, 9.6, and 12.8 for 10, 20, 30, and 40% MF conditions, respectively. Differences were calculated between obtained and expected mean error scores. The hypothesis that the average difference scores over the three complexity groups would be zero was tested. Only the Mean source of variance for the 40% MF group was significant ($P < .01$). Only the 30% MF complexity groups differed significantly among themselves in the extent of their deviations from the predicted value ($P < .05$). This effect was linear as indicated by the orthogonal polynomial, with performance under greater degrees of complexity producing more errors than predicted by probability matching (Table 2).

A tabular analysis on number of left and right button responses revealed the same trend over trials throughout the various treatment conditions.

*Restle's model.*—In Restle's theory (1957), $\delta$ denotes the proportion of relevant cues in the problem. The values of $\delta$ for the three levels of complexity were estimated from the performance of the 0% MF groups. Using these $\delta$s, the value of cue validity, $V$, which would cause each of the 12 MF × Complexity groups to make the mean number of errors observed was computed by Equation 1. Following Restle's procedure (1957, p. 187), the values of $V$ predicted by Equation 2 were found and compared with the computed $V$s. The obtained $V$ for each of the 12 conditions was subtracted from the predicted $V$, and an analysis of variance of these 12 difference scores was performed (Table

TABLE 2

ANALYSES OF VARIANCE FOR DEVIATIONS FROM EXPECTED MEAN ERRORS

| Source | *df* | Misinformation Feedback | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 10% | | 20% | | 30% | | 40% | |
| | | *MS* | *F* | *MS* | *F* | *MS* | *F* | *MS* | *F* |
| Mean | 1 | 12.02 | .59 | .59 | .02 | 62.41 | 2.59 | 201.64 | 9.08** |
| Complexity | 2 | 47.99 | 2.39 | 43.53 | 1.57 | 101.34 | 3.91* | 35.09 | 1.58 |
| Linear | 1 | 60.36 | 3.01 | 60.17 | 2.20 | 170.67 | 6.58 | 20.16 | .91 |
| Quadratic | 1 | 33.89 | 1.69 | 26.89 | .98 | 30.68 | 1.18 | 50.00 | 2.25 |
| Error | 33 | 20.08 | | 27.74 | | 25.59 | | 22.19 | |

* $P < .05$.
** $P < .01$.

TABLE 3

ANALYSIS OF VARIANCE FOR DIFFERENCES BETWEEN OBTAINED AND PREDICTED *V*s

| Source | *df* | *MS* | *F* |
|---|---|---|---|
| Mean | 1 | .002 | 2.00 |
| Cue Validity (*V*) | 3 | .061 | 61.00* |
| Linear | 1 | .18 | 180.00* |
| Quadratic | 1 | .00008 | .08 |
| Cubic | 1 | .0007 | .70 |
| Complexity (C) | 2 | .001 | 1.00 |
| Linear | 1 | .0005 | .50 |
| Quadratic | 1 | .00104 | 1.04 |
| *V* × C (Error) | 6 | .001 | |

* $P < .001$.

3). The main effect of *V* was significant ($P < .001$), indicating some significant discrepancy between predicted and obtained *V*s. This effect was linear as tested by the orthogonal polynomial; deviations becoming greater with increase in Complexity and MF.

One final theoretical analysis concerns the effect of number of irrelevant dimensions. Bourne and Restle (1959) assume that each such dimension is an equal source of irrelevant cues. From this they deduce that $1/\delta$ is a linear function of the number of irrelevant dimensions. Figure 5 is a plot of the reciprocal of estimated $\theta_T$, which here should be proportional to $1/\delta$, in each of the three complexity conditions with all the MF conditions combined. The function is apparently straight-line in form.

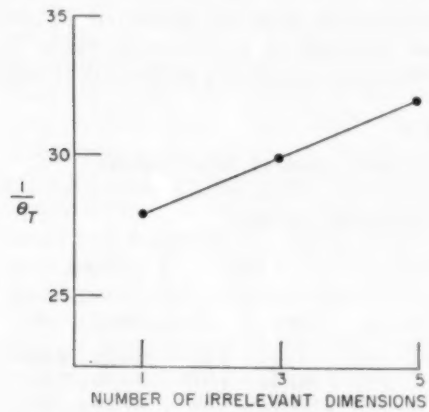



FIG. 5. $1/\theta_T$ as a function of number of irrelevant dimensions. Estimates based on the mean of total errors for each level of complexity over combined degrees of MF.

## DISCUSSION

The results show that the efficiency of concept identification is a decreasing function of MF. This was expected from the results of Bourne and Pendleton (1958) in their study of probability of information feedback. In that study, the omission of feedback on a certain percentage of trials significantly reduced performance. This suggested strongly that MF would have a similar, though greater, effect. Consistent with previous work on effect of task complexity, the present investigation showed that irrelevant information has a quantifiable effect on task difficulty. Furthermore, the effect of MF depends on the level of task complexity, i.e., MF inhibits concept identification more as task complexity increases. It was shown that the slopes of MF × C curves differed significantly on both linear and quadratic components, producing decrement in performance.

The hypotheses concerning the effects of MF were based, in part, on the results of guessing game experiments. Owing to the complexity differences, rate of approach to asymptote ought to be less in the present task, but terminally, *S*s were expected to match probabilities. In the main, these expectations were realized. However, it is important to note that with greatest MF (40%), the number of errors was significantly greater than expected by probability matching. It is regrettable that more trials were not given, since the performance of certain groups might have surpassed probability matching performance, had further opportunity been provided. Trends of this sort were indicated in Fig. 2, 3, and 4.

In this study the sequence of MF was

randomly determined within the 192 patterns. In other studies, e.g., Morin's (1955) the percentage of MF was distributed over blocks of fewer trials, thus providing a more homogeneous distribution of MF. Further investigations of the effects of various distributions of MF would probably shed more light on the validity of $V$. It seems plausible to expect that a homogeneous distribution of MF would render it less effective as an inhibitor of performance. In relation to other studies of MF, present findings must be interpreted with caution. The *S*s' awareness of the existence or possibility of MF has a significant effect on performance as shown by Morin (1955). In this experiment, some indirect suggestion of the presence of MF was made in the instructions. It is probable that task difficulty could be increased by omitting this suggestion.

Some success was met in the theoretical analysis of data based on Restle's (1955, 1957) discrimination model. The $\theta$s, or rates of learning, were somewhat lower in the present data than is the case in simple guessing games (Restle, 1957). Furthermore, $\theta$ tends to decrease as task complexity increases, thus indicating that complexity accounts, at least in part, for the difference. The relationship between $1/\theta_T$ and number of irrelevant dimensions was linear as predicted. Quantitative predictions of errors (Equation 1), however, were grossly divergent from obtained errors for the four degrees of MF, and significant difference between estimated and predicted $V$s were found. This result may stem from two sources. First, there were several unexpected inversions in the obtained error scores, i.e., trends in which errors did not consistently increase with combined increases in MF and Complexity (Fig. 2, 3, and 4). Such inconsistencies mediate against securing correct estimates of certain theoretical parameters, such as $\theta$ and $V$, since these parameters are based on the assumption of additivity of irrelevant cues. Second, it is possible that Restle's concept of cue validity is not appropriate in the analysis of problem solving behavior, since $V$ is based on the function of $\pi$. It is important to note that $\pi$ may be influenced by the differential distributions of MF. Bourne and Restle (1959) have provided evidence that the discrimination model can account for the results of some concept identification experiments, although the present study provides the first test of the cue validity formulation. Present results do not suggest any simple modification of this concept.

## Summary

The primary aim of this experiment was to investigate the effects of misinformative feedback (MF) and task complexity upon concept identification. Each of the 180 *S*s was given 12 blocks of 16 patterns, a total of 192 stimuli. The degrees of MF were 0, 10, 20, 30, and 40% and of complexity, 1, 3, and 5 irrelevant dimensions. There were three problems, each with a different relevant dimension, and equal numbers of men and women *S*s were included in each treatment combination. Essentially, *S*'s task was to learn the correct method for classifying visually presented geometric patterns.

The major findings were: (*a*) Performance improved significantly over trials. (*b*) MF was a significant source of variance in the analysis. The asymptotic mean error scores for the 10, 20, and 30% MF conditions did not significantly differ from the mean predicted values based on a probability matching hypothesis. (*c*) The relationship between errors and Complexity was shown to be linear. Analysis of variance produced a significant MF × Complexity interaction, suggesting an increase in the effect of MF at higher levels of Complexity. (*d*) Sex and Problems effects had no significant influence on performance. (*e*) The relationship between the reciprocal of the estimated proportion of relevant cues and the number of irrelevant dimensions was linear in accordance with Restle's (1955, 1957) learning model. (*f*) Number of errors as predicted by this model was greatly divergent from the obtained errors for all the combined MF × Complexity conditions.

## References

Archer, E. J., Bourne, L. E., Jr., & Brown, F. G. Concept identification as a function of irrelevant information and instructions. *J. exp. Psychol.*, 1955, **49**, 153–164.

BOURNE, L. E., JR. Effects of delay of information feedback and task complexity on the identification of concepts. *J. exp. Psychol.*, 1957, **54**, 201–207.

BOURNE, L. E., JR., & PENDLETON, R. B. Concept identification as a function of completeness and probability of information feedback. *J. exp. Psychol.*, 1958, **56**, 413–420.

BOURNE, L. E., JR., & RESTLE, F. Mathematical theory of concept identification. *Psychol. Rev.*, 1959, **66**, 278–296.

BROWN, F. G., & ARCHER, E. J. Concept identification as a function of task complexity and distribution of practice. *J. exp. Psychol.*, 1956, **52**, 316–321.

ESTES, W. K., & STRAUGHAN, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.

GRANT, D. A., & CURRAN, J. F. Relative difficulty of number, form, and color concepts of a Weigl-type problem using unsystematic number cards. *J. exp. Psychol.*, 1952, **43**, 408–413.

GRANT, D. A., HAKE, H. W., & HORNSETH, J. P. Acquisition and extinction of a verbal conditioned response with differing percentages of reinforcement. *J. exp. Psychol.*, 1951, **42**, 1–5.

HEIDBREDER, E. The attainment of concepts: VI. Exploratory experiments on conceptualization, at perceptual levels. *J. Psychol.*, 1948, **26**, 193–216.

HEIDBREDER, E. The attainment of concepts: VII. Conceptual achievements during card sorting. *J. Psychol.*, 1949, **27**, 3–39.

HUMPHREYS, L. G. Acquisition and extinction of a verbal conditioned response in a situation analogous to conditioning. *J. exp. Psychol.*, 1939, **25**, 294–301.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

MORIN, R. E. Factors influencing rate and extent of learning in the presence of misinformative feedback. *J. exp. Psychol.*, 1955, **49**, 343–351.

MYERS, J. L. On the interaction of two scaled variables. *Psychol. Bull.*, 1959, **56**, 384–391.

RESTLE, F. A theory of discrimination learning. *Psychol. Rev.*, 1955, **62**, 11–19.

RESTLE, F. Theory of selective learning with probable reinforcements. *Psychol. Rev.*, 1957, **64**, 182–191.

(Received June 26, 1959)

*Journal of Experimental Psychology*
Vol. 59, No. 6, 1960

# EYELID CONDITIONING PERFORMANCE UNDER PARTIAL REINFORCEMENT AS A FUNCTION OF UCS INTENSITY [1]

LEONARD E. ROSS[2] AND KENNETH W. SPENCE

*State University of Iowa*

Investigations of the relationship of UCS intensity to performance in eyelid conditioning have consistently employed continuous (100%) reinforcement. Thus, while studies have demonstrated (Spence, 1958) that performance is an increasing function of the intensity of the UCS, evidence is not available concerning the relationship of this variable to eyelid conditioning performance under partial reinforcement schedules. Although there are a number of eyelid conditioning studies in the literature that have used 50% partial reinforcement (Grant & Hake, 1951; Grant & Schipper, 1952; Grant, Schipper, & Ross, 1952; Gynther, 1957; Ross, 1959), a comparison of the performance levels obtained in these experiments as a function of the UCS strength employed is not feasible since in some studies the puff strength was unspecified and, in addition, experimental conditions, which might in themselves account for performance differences, have varied from study to study.

Under 100% reinforcement conditions the relationship between conditioning performance (percentage of CRs, Trials 41–80) and the intensity of the UCS has been found to be a negatively accelerated increasing function (Spence, 1958). The question arises whether or not the relationship under partial reinforcement conditions is of the same general form. The previously cited studies which employed partial reinforcement uniformly found higher levels of conditioning with 100% reinforcement than with 50% partial reinforcement. Since this continuous reinforcement superiority is maintained when comparisons are made in terms of reinforced trials rather than trials alone, the lower level of performance of the partial groups is presumably due to the decremental effects of nonreinforced trials. It has been suggested (Ross, 1959) that these effects are, in part at least, due to factors that might be classified as "set" variables (i.e., under some degree of voluntary control).

Whatever the nature of the factors producing the decrement it seems quite possible that the introduction of nonreinforced trials might result in a relationship between conditioning level and UCS intensity different from that obtained with continuous reinforcement. In the present experiment groups of *Ss* under a random 50% reinforcement schedule were conditioned with different intensities of the UCS in an effort to provide information concerning the function relating conditioning performance under partial reinforcement conditions to the strength of the UCS. In addition, one group was conditioned with continuous reinforcement to supplement existing 100% data so that 50%–100% comparisons could be made at an additional UCS intensity value.

[1] This study is part of a project concerned with the influence of motivation on performance in learning under Contract N9 onr-93802, Project NR 154-107 between the State University of Iowa and the Office of Naval Research. Robert S. Blanchard helped in the collecting of the data.

[2] Now at the University of Wisconsin.

## METHOD

*Subjects.*—The *Ss* were 115 women and 58 men from an introductory psychology course. Of these, 10 men and 6 women were discarded for exceeding the criterion used to exclude voluntary responders (Spence & Ross, 1959), while 1 woman was excluded from the data for giving CRs to test trial presentations of the CS alone. Six additional *Ss* were discarded, three due to *E* error and three due to apparatus failure. The remaining 150 *Ss* were assigned at random to five groups except that the proportion of women to men was fixed at a ratio of 7 to 3.

*Apparatus and method of recording.*—The apparatus and method of recording which involved a potentiometer–polygraph system, has been described elsewhere (Spence & Taylor, 1951). The CS was an increase in brightness of a 2¼-in. diameter circular milk glass disk from less than .004 apparent ft-c to .32 apparent ft-c. The UCS was a 50-msec. air puff delivered through a .062-in. diameter orifice to the right eye.

*Conditioning procedure.*—Each *S* was given instructions to blink once to the ready signal and then to look at the disk on the wall, remaining as relaxed as possible, until the light went out. Following the reading of the instructions three test trial presentations of the CS alone and one presentation of the UCS alone were given. Immediately following this preliminary procedure each *S* received 140 trials at intertrial intervals of 15, 20, and 25 sec., averaging 20 sec., and arranged according to a fixed schedule. The word "ready" preceded each presentation of the CS by 2, 3, or 4 sec. according to a prearranged schedule.

Four groups of 30 *Ss* were conditioned with a 50% reinforcement schedule consisting of half reinforced and half nonreinforced trials, presented in a prearranged sequence within which neither type of trial occurred more than twice in succession. A rei forced trial involved a 500-msec. CS-UCS interval while on nonreinforced trials the UCS followed the CS onset at 2500 msec. Each group was conditioned with a different intensity of UCS. The four values used were .33, .6, 1.0, and 4.0 psi. An additional group of 30 *Ss* was given 100 trials with continuous reinforcement. The UCS for this group was a .33 psi air puff.

## RESULTS AND DISCUSSION

Frequency curves of conditioning by blocks of 10 reinforced trials for the four groups receiving partial reinforcement are shown in Fig. 1. Inspection of the figure reveals that only the 4.0-lb. psi group showed an appreciable performance increase, the other three groups remaining under the 20% CR level throughout the course of conditioning. Although the 1.0-lb. group was superior to the .33-lb. group the performance level of the .6-lb. group was slightly lower than that of the .33-lb. group.

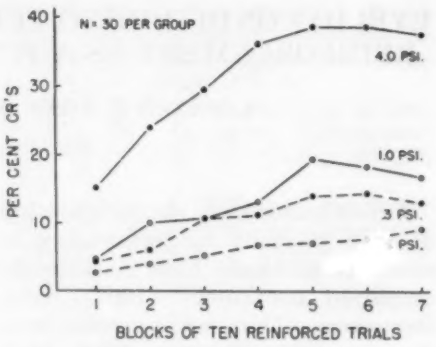



FIG. 1. Mean percentage of CRs in successive blocks of 10 reinforced trials.

To statistically evaluate these differences a Kruskal-Wallis (1952) *H* test was employed with the scores for the four groups over conditioning Trials 51–70 (Trials 100–140). The resulting *H* of 20.64 was significant at the .001 level. Subsequently, Mann-Whitney *U* tests were used to test the observed differences for

TABLE 1

MANN-WHITNEY *U* AND THE CORRESPONDING *P* VALUES FOR ALL POSSIBLE COMPARISONS OF THE GROUPS ON TRIALS 41–70

| Group Comparisons | *U* | *P* |
|---|---|---|
| .33– .6 | .75 | .453 |
| .33–1.0 | .90 | .368 |
| .33–4.0 | 3.47 | .0006 |
| .6 –1.0 | 1.65 | .099 |
| .6 –4.0 | 4.09 | .0001 |
| 1.0– 4.0 | 2.80 | .005 |

all possible pairs of puff strength values. Table 1 presents the *U* and corresponding *P* value for each of the comparisons. It should be noted that while the 4.0-lb. group differed significantly from each of the other groups, none of the other comparisons reached the .05 level of significance. In view of these statistical results the differences among the three low puff strength groups must be considered to be fortuitous. It would appear that with UCS values of 1.0 lb. and less a low performance level is obtained and that within this range there is little, if any, positive relationship between UCS intensity and conditioning performance.

In order to compare the function relating conditioning performance to UCS intensity for 50% partial and continuous reinforcement conditions, the data of a number of published studies from the Iowa laboratory were used. Except for the intensity of the UCS these studies were highly comparable to the present study in terms of apparatus, conditioning procedure, and experimental conditions (i.e., CS, CS-UCS interval, intertrial interval, etc.). Since the use of a frequency measure imposes a ceiling on performance (100%), and a number of *Ss* in the strong puff groups had reached this ceiling, the comparisons were made in terms of the conditioning performance of homogeneous *Ss*, i.e., the middle 40% of each group. Thus the *Ss* from each group were rank ordered on the basis of the total number of responses made during the first 70 conditioning trials and the mean percentage of CRs made by the middle 40% of each distribution was computed. Because of varying ratios of men to women this procedure was followed for each sex separately and the mean percentage values obtained were averaged for each experiment. The results obtained are presented in Fig. 2. For the continuous reinforcement function the .33-lb. point was taken from the 100% reinforcement group in the present

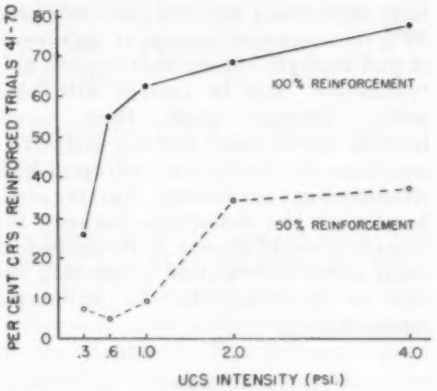



FIG 2. Relationship between percentage of CRs and UCS intensity for groups of homogeneous *Ss* (see text) receiving partial and continuous reinforcement.

experiment while the other points, from .6 to 4.0 lb., respectively, were obtained from studies by Runquist and Ross (1958), Runquist and Spence (1959), Ross (1959), and Spence and Ross (1959). All of the points on the partial reinforcement function were taken from the present experiment except for the 2.0-lb. point which was obtained from an earlier study (Ross, 1959).

From an inspection of Fig. 2 it is evident that the shapes of the functions are quite different for the two reinforcement schedules. With continuous reinforcement the function is negatively accelerated with small increases in puff strength at low absolute levels resulting in large increases in performance. In contrast, under 50% reinforcement the function is apparently S-shaped with little performance change over the low range of UCS intensity values.

Just why such different results should be obtained with the introduction of nonreinforced trials into the conditioning situation is not clear. However, it is possible that, if the decrement found with 50% reinforcement is to some degree related to "set" factors, inhibition of performance is more readily accomplished under conditions of low puff strengths. Thus, increases in puff strengths over this range would be ineffective in increasing performance. The

large differences between the 100% and 50% reinforcement groups at high levels of puff strength require that considerable "inhibition" still be present with such puffs. Stronger puffs, then, would increase performance but not completely overcome the "inhibition" effects of 50% reinforcement conditions. Further clarification of the differences between the functions would appear to be dependent upon added information concerning the role of inhibitory factors in partial reinforcement.

## Summary

Groups of *Ss* under a random 50% reinforcement schedule were conditioned with different intensities of UCS (.33, .6, 1.0, 4.0 psi) in an effort to provide information concerning the function relating conditioning performance under partial reinforcement conditions to the strength of the UCS. At UCS values of 1.0 lb. and less, a low performance level was obtained, with little indication of a positive relationship between performance and UCS intensity. A 4.0-lb. UCS resulted in performance significantly higher than that of the other groups.

In order to compare the function relating conditioning to UCS strength obtained for partial reinforcement in the present experiment with one for continuous reinforcement, the data from a number of previously published studies were analyzed and a similar function determined for 100% reinforcement. The functions differed in shape, the partial reinforcement function being S-shaped, while that for continuous reinforcement appeared to be negatively accelerated throughout.

## References

Grant, D. A., & Hake, H. W. Dark adaptation and the Humphries random reinforcement phenomenon in human eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 417–423.

Grant, D. A., & Schipper, L. M. The acquisition and extinction of conditioned eyelid responses as a function of the percentage of fixed ratio random reinforcement. *J. exp. Psychol.*, 1952, **43**, 313–320.

Grant, D. A., Schipper, L. M., & Ross, B. M. Effect of intertrial interval during acquisition on extinction of the conditioned eyelid response following partial reinforcement. *J. exp. Psychol.*, 1952, **44**, 203–210.

Gynther, M. D. Differential eyelid conditioning as a function of stimulus similarity and strength of response to the CS. *J. exp. Psychol.*, 1957, **53**, 408–416.

Kruskal, W. H., & Wallis, W. A. Use of ranks in one-criterion variance analysis. *J. Amer. Stat. Ass.*, 1952, **47**, 583–621.

Ross, L. E. The decremental effects of partial reinforcement during acquisition of the conditioned eyelid response. *J. exp. Psychol.*, 1959, **57**, 74–82.

Runquist, W. N., & Ross, L. E. A preliminary investigation of the relationship between GSR, heart rate change, and eyelid conditioning. *Proc. Iowa Acad. Sci.*, 1958, **65**, 398–404.

Runquist, W. N., & Spence, K. W. Performance in eyelid conditioning as a function of UCS duration. *J. exp. Psychol.*, 1959, **57**, 249–252.

Spence, K. W. A theory of emotionally based drive (*D*) and its relation to performance in simple learning situations. *Amer. Psychologist*, 1958, **13**, 131–141.

Spence, K. W., & Taylor, J. A. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.

Spence, K. W., & Ross, L. E. A methodological study of the form and latency of eyelid responses in conditioning. *J. exp. Psychol.*, 1959, **58**, 376–381.

(Received June 26, 1959)

*Journal of Experimental Psychology*
Vol. 59, No. 6, 1960

# EFFECT OF TRIPLET AND QUADRUPLICATE LOCATION IN VERBAL MAZE PATTERNS UPON SERIAL POSITION ERRORS[1]

GEDIMINAS NAMIKAS, CHARLES P. THOMPSON, AND W. J. BROGDEN

*University of Wisconsin*

The effect of change in pattern of the 16-unit, 4-choice verbal maze of Brogden and Schmidt (1954) upon its acquisition and form of the serial position error curve has been studied by Thompson, Voss, and Brogden (1957) and by Ernst, Hoffeld, Seidenstein, and Brogden (1960). Alteration of the standard maze pattern by the introduction of a doublet (two successive like response items) or a split-doublet (two identical response items separated by a different response item) changes the form of the serial position error curve but does not affect speed of acquisition. A doublet in each of five different locations of the maze pattern produces an increase in error at the position of the second item. A split-doublet in each of four different locations (except for Positions 2–4) produces an increase in error at the third position (second like item). With the one exception for the split-doublet, there is no differential effect of doublet or split-doublet location in the maze pattern.

The doublet effect is maximum for the total error measure early in acquisition and decreases in magnitude thereafter. For the first error measure, there is no doublet effect over all trials. For the first one-quarter of trials the doublet effect is present as an increase of error at the position of the second item, but for the last three-quarters of trials the effect is reversed and occurs as a decrease in error at the second position of the doublet. The presence of a doublet in the maze pattern thus has an inhibitory effect upon acquisition during the discovery phase and this is a function primarily of the guessing habits of *S*s, particularly the tendency not to repeat a response correct at the previous position. Once discovery is completed, the identity of the doublet items facilitates further acquisition of the second item of the doublet.

The split-doublet phenomenon is apparently different from that for the doublet. There is no evidence of difference between the total error and first error measures over all trials or for stages of acquisition, nor of a relation between guessing habits of *S*s and the split-doublet effect. These results suggest that there may be no consistent effect of a pattern dimension upon the serial position error curve and that variables representative of the dimension such as the doublet and split-doublet have specific and independent effects. Within the parameters of the 16-unit, 4-choice verbal maze pattern, with each choice represented four times, the triplet (three successive like responses) and the quadruplicate (four successive like responses) are the only other sources of pattern variation that can be included in a dimension that also includes the doublet and split-doublet. The effects of the triplet

[1] This research was supported in part by grants from the National Science Foundation and the Research Committee of the Graduate School from funds provided by the Wisconsin Alumni Research Foundation.

and quadruplicate upon the form of the serial position error curve were hypothesized to be different from those of the doublet and split-doublet, to be different from each other, to involve both increases and decreases in error as a function of position within the triplet or quadruplicate sequence, and to show no difference in character as a function of location within the maze pattern. It was hypothesized further that the introduction of a triplet or quadruplicate would increase the speed of acquisition over that for the standard maze pattern. This paper presents the results of two experiments, one in which the triplet and its location represent the independent variable and the other for which the independent variable is the quadruplicate and its location in the maze pattern.[2] These two experiments follow closely the design of the earlier studies on the doublet and split-doublet in an attempt to test the hypotheses presented above.

## Procedure

*Experiment I.*—The four experimental groups acquired a modification of the standard maze pattern which includes a single triplet located at Maze Units 3, 4, and 5, at 6, 7, and 8, at 9, 10, and 11, and at 12, 13, and 14, respectively. Each modified pattern was altered not only to locate the triplet at the appropriate positions, but also to exclude doublet, split-doublet, or other noticeable sequences in the maze. Each pattern was permuted to obtain four subpatterns and thus provide for each number (10, 20, 30, or 40) to be represented once at each of the 16 positions. Data for the control pattern is that reported by Ernst et al. (1960).

Of 20 *Ss* in each of the four experimental groups, five learned each one of the four subpatterns. All of the *Ss* were treated by a single *E*. The *Ss* were volunteer students from the classes in elementary psychology at the University of Wisconsin. Assignment of *Ss* to experimental conditions was random within each replication.

The instructions of Thompson et al. (1957) were read to *S*. Responses of 10, 20, 30, or 40 were made by *S* with *E* signaling the correct response by saying "Forward." This procedure continued until *S* completed one errorless trial. Each response at each unit was recorded by *E* for each trial. A stop watch was used to measure the time of each trial and the 1-min. rest between trials. Other conditions were comparable to those of Thompson et al. (1957).

*Experiment II.*—The patterns learned by the three experimental groups contained a single quadruplicate located at Maze Units 3, 4, 5, and 6, at 7, 8, 9, and 10, and at 11, 12, 13, and 14, respectively. Each pattern represents a limited modification of the standard pattern to provide a quadruplicate at the desired locus without other noticeable pattern sequences elsewhere in the maze. In all other respects, the procedure for Exp. II was identical to that of Exp. I.

## Results

*Experiment I.*—Analyses of variance of the data for time and trials for the four experimental and one control group showed no significant effect for maze pattern, maze pattern subgroups, or the interaction of these two variables. Since the assumption of homogeneity of variance was found not tenable for the error data, the 1% level of significance was selected in lieu of the 5% level for all analyses. No source from the between-*Ss* variation was significant. Further analysis of the significant interaction of serial position by maze pattern was based upon several conversions of the data.

Total error to the criterional trial at each serial position was converted to percentage of total error for each *S* in each group (McCrary & Hunter, 1953). The group mean of the percentage scores for *Ss* of each experimental group and the combined control group of Ernst et al. (1960) are presented in Fig. 1 by

[2] Thompson collaborated with the senior author on the triplet experiment and Namikas on the experiment with the quadruplicate.

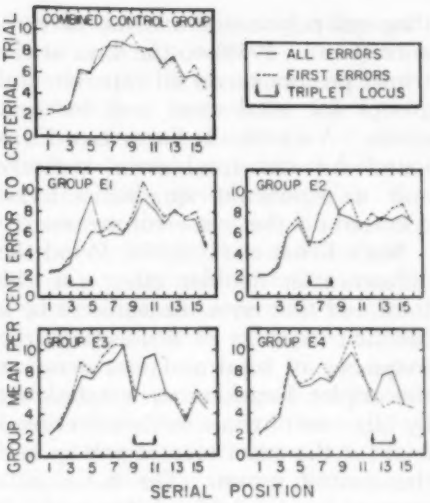



FIG. 1. Serial position curves in terms of percentage of total errors and percentage of first errors as a function of triplet locus.

the solid line curves. There are substantial differences in the form of the curves for the experimental groups compared with each other and with the control group. These differences are clearly shown by the solid line curves in the left half of Fig. 2. The data for these curves were obtained by taking the difference in percentage of error for each *S* and the mean percentage error for the combined control group at each of the 16 serial positions and computing the mean of these differences for each of the four experimental groups. For each group there is an increase in error for the third position of the triplet, and except for Group $E_1$, there is a decrease in error for the first position. Analysis of variance of these data shows in each case a significant interaction of serial position and maze pattern and thus establishes significant differences in the form of the difference curves between the experimental groups and the control group and between the four experimental groups.

The broken line curves of Fig. 1 are the group mean percentage of first errors over all trials to the criteria and the broken line curves of Fig. 2 are the difference curves between each experimental group and the control group. The curves in the right half of Fig. 2 give the difference between percentage total error and percentage first error for the experimental groups. The only substantial differences between the all error and first error measures occur at Positions 2 and 3 of the triplet. Analyses of variance of the two sets of first error data and of the differences between percentage total error and percentage first error show in each case a significant serial position by maze pattern interaction.

The *t* test was used to establish fiducial limits at the 1% level of significance for the three sets of mean difference data shown in Fig. 2.

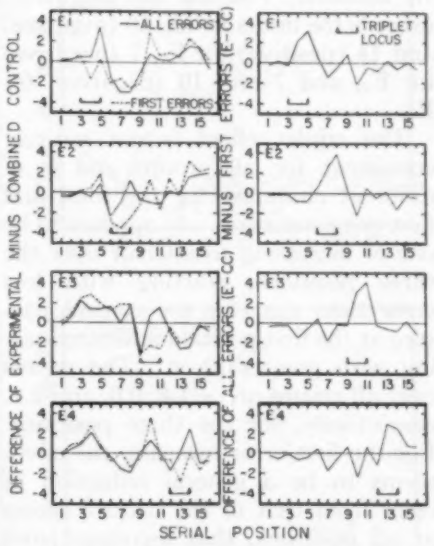



FIG. 2. Curves of differences between serial position curves for experimental minus control groups for percentage of total and first errors (left side) and curves of differences between these differences for each experimental group (right side).

These limits are ±3.1 for all errors, ±2.9 for first errors, and ±1.6 for the difference between all errors and first errors. For the all error measure, the differences from the control group that equal or exceed the fiducial limits for triplet locus are the first triplet item (negative) for Group $E_3$, and the third triplet item (positive) for Group $E_1$. No positions outside of the triplet locus exceed the limits.

For the first error measure, the first item of the triplet exceeds the negative limit for Groups $E_2$ and $E_3$, and the second item of the triplet exceeds the negative limit for Groups $E_2$ and $E_4$. Other significant differences are Positions 10 (positive) for $E_1$, 14 (positive) for $E_2$, 4 (positive) for $E_3$, and 10 (positive) for $E_4$.

For the difference between total error and first error, the second and third items of the triplet exceed the positive fiducial limit for every group. In addition, Position 10 (negative) exceeds the limits for $E_1$, 10 (negative) and 14 (positive) for $E_2$, 7 (negative) for $E_3$, and 7 and 10 (negative) for $E_4$.

The triplet effect is not entirely consistent for all groups and is of different character for the total and first error measures. It appears to be one of increasing total error over the three positions, starting with less error than that for the control pattern at the first position and terminating with greater error. The means over all groups are −1.8, 0.0, and 2.1, respectively, for the three positions. For the first error measure, the effect seems to be a general reduction in error from that of the control group at all positions, that increases from the first to second item and decreases from the second to the third. The means over all groups are −1.5, −2.6, and −0.8, respectively. These apparent trends were tested by orthoganol polynomial analysis of variance (Grant, 1956) on the data at the triplet positions over all experimental groups for total error and for first errors. A significant linear trend was found for the total error measure and a significant quadratic trend occurred for the first error measure.

Since Ernst et al. (1960) found the difference in doublet effect for the total and first error measures to be a function of stage of acquisition, the frequency of total and first error at the triplet locations were tabulated by quarters of trials to the criterional trial for the experimental groups and the control group. The E-CC differences were obtained for each *S* of the experimental groups and the over-all group means for total and first error measures computed for each triplet position. All differences for both measures are positive for the first quarter of trials and negative for each one of the last three quarters of trials except for the total error measure for the third triplet item for the second quarter. The linear trend for total errors and the quadratic trend for first errors do not occur consistently for each quarter of trials, but for both measures the effect appears to be maximum for the first quarter and to decline progressively over the remaining three quarters. Each measure shows its characteristic trend, but early in learning the trend occurs over errors greater than those for the control group and later in acquisition, over errors less than the control group. No statistical tests were made on these data.

*Experiment II.*—Analyses of variance of the data for time and trials for the three experimental and one control group showed no significant effect for maze pattern, maze pattern subgroups or the interaction. Since

the assumption of homogeneity of variance was untenable for the error data, the 1% level of significance was selected in lieu of the 5% level for all analyses. Treatment of the error data for this experiment follows that described already for Exp. I, with comparable results. The interaction of Serial position × Maze pattern is significant for total errors, for percentage total errors, for percentage first errors, for percentage total errors experimental minus control, and for percentage first errors experimental minus control. The serial position error curves for percentage of all error and for percentage of first error for the control group and the three experimental groups are presented in Fig. 3. The effect of the quadruplicate appears to be one of general reduction of error. This is shown by the curves of difference between the experimental and control groups for percentage total errors and percentage first errors which are presented in the left half of Fig. 4. The curves in the right half of Fig. 4 are the differences between percentage total errors and percentage first errors for each experimental group. The curves for all error and for first error follow each other closely except that the first error curves for $E_1$ and $E_2$ appear to show a greater reduction in error for the quadruplicate locus than do those for total error. The curves of difference between total and first errors appear to reflect these differences primarily. Analysis of variance of the difference data between total and first errors reveals a significant interaction of serial position and maze pattern.

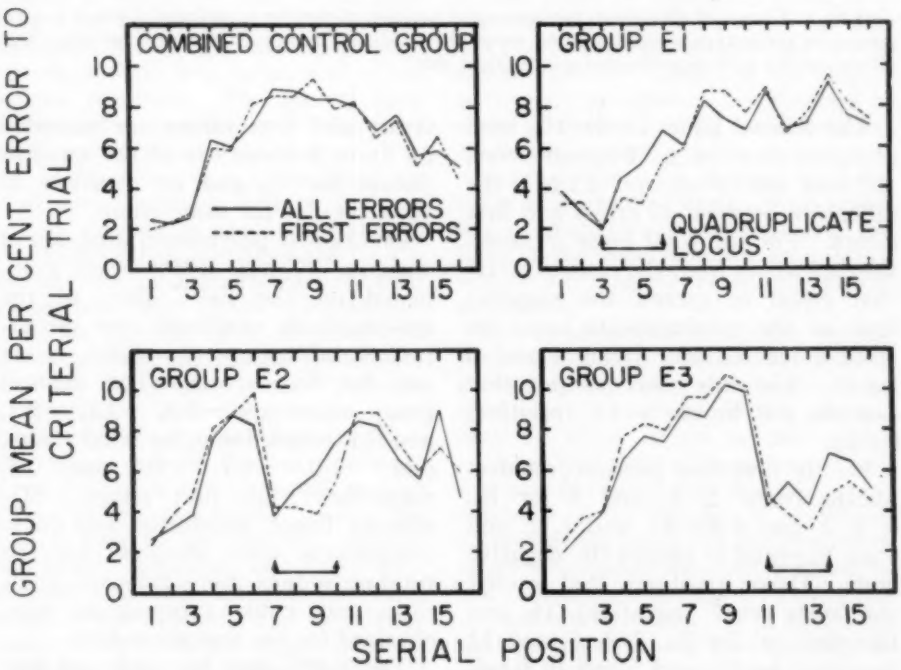



FIG. 3. Serial position curves in terms of percentage of total errors and percentage of first errors as a function of quadruplicate locus.

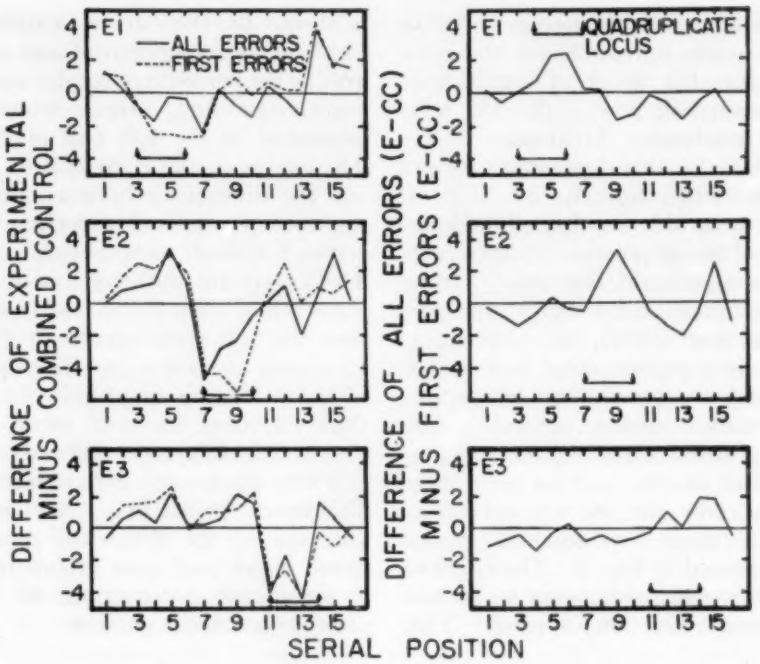



FIG. 4. Curves of differences between serial position curves for experimental minus control groups for percentage of total and first errors (left side) and curves of differences between these differences for each experimental group (right side).

The fiducial limits at the 1% level of significance are ±3.6 for all errors, ±2.0 for first errors, and ±2.6 for the difference between all errors and first errors. For the total error measure, the differences from the control group that equal or exceed the negative limit at the quadruplicate locus are Item 1 for $E_2$, and Items 1 and 3 for $E_3$. The only other position that exceeds the limits is 14 (positive) for $E_1$.

For the first error measure, quadruplicate Items 2, 3, and 4 for $E_1$, 1, 2, 3, and 4 for $E_2$, and 1, 2, and 3 for $E_3$ equal or exceed the negative limit. Other positions that exceed the limits are 7 (negative), 14, and 16 (positive) for $E_1$, 3, 4, 5, and 12 (positive) for $E_2$, and 5 and 10 (positive) for $E_3$. The fiducial limits for the curves of difference between all errors and first errors are exceeded by Item 3 (positive) of the quadruplicate for $E_2$ and by Position 15 (positive) for the same group.

Orthogonal polynomial analyses of variance (Grant, 1956) were completed on the E-CC data of the quadruplicate positions over all experimental groups for total errors and for first errors. The over-all group means are −3.3, −2.1, −2.3, and 0.1, respectively, for total errors, and −1.0, −0.7, −0.5 and 0.5, respectively, for first errors. Significant linear, quadratic, and cubic components were obtained for the total error data, and significant quadratic and cubic components were obtained for the first error data.

The E-CC data for total and first errors at the positions of the quadruplicate were each tabulated by quar-

ters of trials to the criterial trial. All means over groups show less error than the control group except the last three positions of the quadruplicate for the first quarter of trials for the total error measure. These latter means are progressively positive in magnitude. The trend over the four positions of the quadruplicate is more apparent during the first two quarters of trials for both measures and is more evident with the total error measure than with the first error measure. No statistical tests were made of these data.

## Discussion

The two experiments of this study establish a triplet effect and a quadruplicate effect on the form of the serial position error curve. Locus of the triplet and the quadruplicate does not appear to be differential, except that location of the pattern early in the maze (3, 4, 5 for the triplet and 3, 4, 5, 6 for the quadruplicate) produces a less clear-cut effect than does location of either pattern elsewhere. The general form of each effect is different for the total error and first error measures. For the triplet, there is a linear trend for total error, starting with a decrease in error relative to the control pattern for the first item, no difference for the second item, and an increase in error for the final item. There is a quadratic trend for the first error measure, with less error than the control pattern for each of the triplet items. For the quadruplicate, both the total error and first error measures show less error than the control pattern for the first three positions and a zero or small positive difference for the fourth item. The general trend for the total error measure has linear, quadratic, and cubic components and that for the first error measure has quadratic and cubic components. For both the triplet and quadruplicate, there are other differences in form of the serial position error curve that may not be accounted for by the location of either pattern or by the interchanges of response items necessary to pattern formation.

The triplet and quadruplicate effects are not only of different character from each other but differ also from the doublet and the split-doublet effects (Ernst et al., 1960). The doublet effect stands apart from the other three pattern effects in that it is in part a function of the guessing habits of *Ss* and the effect measured by first errors is diphasic. Although the quadruplicate effect as measured by total errors shifts from positive error relative to the control group for the first quarter of trials to negative for each of the last quarters, the trend over both triplet and quadruplicate positions is relatively stable over the course of acquisition. The differences in the triplet and quadruplicate effects between total and first errors appear to be a function of the differences in the nature of these measures.

It does appear that as the number of like items is increased, there is a progressive decrease in error at the locus involved, going from an initial increase in error relative to the control pattern for the doublet to a decrease in error maximum for the quadruplicate. All show a trend within the pattern of a relative increase in error from the initial to succeeding items. Each represents a characteristic modification in the form of the serial position error curve without affecting the over-all measures of acquisition.

The quadruplicate represents the upper limit of a pattern dimension involving succeeding like items with fixed parameters of a 16-unit maze of four choices, each of which occurs as the correct response four times. Longer mazes with a greater number of choices would permit extension of the variable of number of succeeding like items and should produce larger decrements in error relative to the control pattern than that produced by the quadruplicate. It is likely also that the pattern effects beyond the quadruplicate in longer mazes will produce an over-all decrease in measures of acquisition relative to the control

pattern and differential effects as a function of locus. The results from the triplet and quadruplicate studies suggest that their effects are reduced for the early positions of the maze, as is the case also for the split-doublet.

## Summary

Two experiments were designed to test respectively the effect of location of a triplet and a quadruplicate on the form of the serial position error curve from the acquisition of a 4-choice, 16-unit verbal maze. Significant differences were obtained in both experiments in the form of the curves between each experimental group and the control group, and between experimental groups. The effect was less clear for the locus early in the pattern (3, 4, 5 for triplet and 3, 4, 5, 6 for quadruplicate) than for the other locations. The triplet effect is a linear trend over the three positions for the total error measure, with over-all mean differences relative to the control group of −1.8, 0.0, and 2.1. For the first error measure, the respective means are −1.5, −2.6, and −0.8, with a significant quadratic trend. The quadruplicate effect is represented by mean differences over the four positions of −3.3, −2.1, −2.3, and 0.1 for total errors and −1.0, −0.7, −0.5, and 0.5 for first errors. For the former, there are linear, quadratic, and cubic components, and for the latter, quadratic and cubic components. No relation was found between these phenomena and the guessing habits of *Ss*. The phenomena are relatively stable over successive stages of acquisition, and there are no marked differences between total and first errors at the different stages of acquisition. The results are discussed relative to the doublet and split-doublet effects and in relation to a general pattern dimension and its effect upon the serial position error curve.

## References

Brogden, W. J., & Schmidt, R. E. Effect of number of choices per unit of a verbal maze on learning and serial position errors. *J. exp. Psychol.*, 1954, **47**, 235–240.

Ernst, R. L., Hoffeld, D. R., Seidenstein, S., & Brogden, W. J. Relation of serial position errors to doublet and split-doublet location in verbal maze pattern. *J. exp. Psychol.*, 1960, **59**, 94–103.

Grant, D. A. Analysis-of-variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53**, 141–154.

McCrary, J. W., Jr., & Hunter, W. S. Serial position curves in verbal learning. *Science*, 1953, **117**, 131–134.

Thompson, R. F., Voss, J. F., & Brogden, W. J. Effect of pattern variation upon verbal maze learning. *J. exp. Psychol.*, 1957, **54**, 253–258.

(Received June 26, 1959)

*Journal of Experimental Psychology*
Vol. 59, No. 6, 1960

# EFFECT OF CONCURRENT RESPONSES ON THE EVOCATION AND GENERALIZATION OF THE CONDITIONED EYEBLINK[1]

G. ROBERT GRICE[2] AND JOHN D. DAVIS[3]

*University of Illinois*

Recently, the authors (Grice & Davis, 1958) have presented evidence for a mediated generalization effect in human eyelid conditioning. The experiments involved a differential conditioning paradigm in which the middle of three tones served as the CS. The *S*s also made concurrent manual reaction-time responses to the three tones. It was found that significantly more CRs were made to the negative stimulus to which the manual response was the same as the CS than to the one to which the manual response was different. This asymmetrical generalization gradient was interpreted as being in accord with a theory of mediated stimulus generalization, and the advantage was claimed for the design that the "mediating reactions" were overt and subject to experimental control. The results of these experiments were seriously complicated, however, by a direct facilitative effect of the manual reactions on the eyeblink response. This effect appeared to be more powerful than the mediational one and introduced some danger in the interpretation of the experiments. For example, if the manual reactions to the "Same Response" tone should be faster than to the "Different Response" tone the asymmetrical generalization could be an artifactual result of differential facilitation, rather than an associative, mediational effect.

The first purpose of the present experiments is to investigate the relation between these two effects by means of an analysis of the relation between the latencies of the manual reactions and the occurrence of eyeblinks. The second purpose is to compare the mediational effect in a situation where the CS and the generalized stimuli are in the same sense modality with one in which they are different. Theoretically, while primary stimulus generalization would be substantially greater within one modality, the amount of mediated generalization should be independent of the fact that the CS and generalized stimuli are in different sensory dimensions.

## METHOD

### *Apparatus*

The experimental situation was the same as previously described (Grice & Davis, 1958) with the exception of certain improvements in instrumentation. The tones were presented by means of Permoflux PDR-8 earphones rather than by means of a loudspeaker. The same three tones, 240, 850, and 1900 cps, were presented at 70 db loudness level with reference to an average threshold for 1000 cps. They were switched on and off by means of a Grayson-Stadler electronic switch, which was set with a 30 msec. rise and decay time. A continuous "white" masking noise of 20 db sensation level was also presented. The UCS was a puff of air of 1 lb./sq. in. pressure, and the CS-UCS interval was .5 sec. Programming equipment made possible the automatic presentation of

[1] This research was supported by a grant from the National Science Foundation.

[2] In the absence of the junior author, the senior author was entirely responsible for the preparation of the present report.

[3] Now at the American University of Beirut.

the entire experimental session of 100 trials, with attendant reduction of *E* error and general increase in precision.

The manual reaction latencies were recorded on the oscillograph tape with the CRs and were read to the nearest 8 msec. Reactions in excess of 640 msec. were recorded at that value.

### *Procedure*

*Experiment I.*—With the exception of the apparatus differences noted, the procedure was the same as in Exp. I of the previous paper. The same counterbalancing of the high and low tones and right and left hand responses was employed. The middle tone always served as the CS, the air puff never being paired with either of the other tones. Criteria for scoring CRs and elimination of voluntary responders were the same, although no voluntary responders were actually eliminated in either experiment.

The *S*s were 12 men and 12 women students who served as part of a requirement for a course in general psychology. Eight additional *S*s who made fewer than 12 CRs were eliminated. The number of errors in the reaction-time task ranged from 0 to 6 with a mean of 2.0 and a median of 1.85.

*Experiment II.*—Procedure was the same as in Exp. I, except that the CS was a frosted 7-w. bulb placed ¾ in. behind the 1-in. milk glass disk used as a fixation point in the previous experiments. The negative stimuli were the same 240- and 1900-cps tones used in Exp. I.

The *S*s were 12 men and 12 women from the same population as Exp. I. Three *S*s who made fewer than 12 CRs were eliminated. Errors on the reaction-time task ranged from 0 to 5, with a mean of 2.21 and a median of 2.0.

## Results

All data presented are for Trials 41–100, which included 20 presentations of each of the three stimuli. The mean number of CRs to each of the three stimuli in the two experiments are presented in Table 1. The predicted asymmetry was obtained in both experiments, there being more responses to the negative stimulus to which the manual reaction was the same as to the CS than to the one for which it was different. Experiment I replicates the results of Exp. I in the previous paper except that the above-100% generalization to the Same Response stimulus found in that experiment was not obtained.

Mean latencies of the manual responses are presented in Table 2. For Exp. I, the reaction times were somewhat faster to the Same Response stimulus. There was little difference between the reaction times to these two stimuli in Exp. II.

In order to more fully evaluate the effects of the mediational and facilitative effects, point biserial correlation coefficients were computed over trials between reaction time and the occurrence or nonoccurrence of CRs for each *S* separately. The means of these correlations, based on the *z* transformation, are presented in Table 2. The number of CRs to each stimu-

TABLE 1

MEANS OF UNADJUSTED AND ADJUSTED NUMBER OF CRs

| Stimulus | Experiment I | | | | Experiment II | | | |
|---|---|---|---|---|---|---|---|---|
| | Unadjusted | | Adjusted | | Unadjusted | | Adjusted | |
| | Mean | *SD* | Mean | *SD* | Mean | *SD* | Mean | *SD* |
| Same Resp. | 13.25 | 4.18 | 13.01 | 3.69 | 8.33 | 5.56 | 8.91 | 5.26 |
| CS | 14.46 | 3.55 | 14.15 | 4.08 | 16.00 | 2.79 | 14.73 | 3.06 |
| Diff. Resp. | 11.00 | 5.77 | 11.56 | 5.10 | 5.96 | 6.19 | 6.66 | 5.87 |

TABLE 2

REACTION-TIME MEANS AND AVERAGE POINT BISERIAL CORRELATIONS BETWEEN REACTION TIME AND CR OCCURRENCE

| Stimulus | Reaction Time in Msec. | | | | $r_{Pbis}$ | | | |
|---|---|---|---|---|---|---|---|---|
| | Exp. I | | Exp. II | | Exp. I | | Exp. II | |
| | Mean[a] | *SD* | Mean[a] | *SD* | Mean[b] | Range | Mean[b] | Range |
| Same Resp. | 508 | 73 | 500 | 59 | −.63 | +.30 to −.89 | −.58 | −.11 to −.86 |
| CS | 509 | 78 | 454 | 36 | −.59 | −.08 to −.81 | −.63 | −.36 to −.87 |
| Diff. Resp. | 526 | 61 | 503 | 64 | −.46 | +.04 to −.86 | −.29 | +.32 to −.78 |

[a] Mean of *S* means.
[b] Obtained by *s* transformation.

lus was then adjusted for each *S* separately by means of his own "within stimulus" regression coefficient. This leaves the *S* totals unchanged, but adjusts the number of CRs to each stimulus for each *S* for reaction time, and, presumably, for within-*S* variation in the facilitative effect. Means of the adjusted measures are presented in Table 1. It may be noted that the asymmetry of generalization was preserved for both experiments.

Separate analyses of variance of the adjusted numbers of CRs for the two experiments are presented in Table 3. While the Generalization effect is significant for both experiments, the appropriate test for the predicted asymmetry of the mediation hypothesis is a *t* test, based on the within-*S*s error, of the differences between the Same Response and Different Response negative stimuli. This test was significant for Exp. II, ($P < .02$) and of borderline significance for Exp. I. ($P < .05$, one tail).

TABLE 3

ANALYSES OF VARIANCE OF ADJUSTED NUMBER OF CONDITIONED RESPONSE

| Source | *df* | *F* | |
|---|---|---|---|
| | | Exp. I | Exp. II |
| Between *S*s | 23 | | |
| Tone arrangement (T) | 1 | .0005 | .12 |
| Hand arrangement (H) | 1 | .35 | .88 |
| T × H | 1 | 4.84 | .002 |
| Error (b) | 20 | (37.85)[a] | (62.52)[a] |
| Within *S*s | 48 | | |
| Generalization (G) | 2 | 5.50* | 48.03** |
| G × T | 2 | 4.25* | .40 |
| G × H | 2 | .97 | 1.93 |
| G × H × T | 2 | .04 | .46 |
| Error (w) | 40 | (7.38)[a] | (8.66)[a] |
| Total | 71 | | |

[a] Numbers in parentheses are error mean squares.
* $P < .025$.
** $P < .001$.

The negative stimuli were identical in the two experiments. Table 4 is the summary of a combined analysis of the adjusted numbers of CRs to the two negative tones only. The significant effect for Experiments indicates the obvious finding of greater primary generalization from a tone to other tones than from a light to tones. Since responses to the CS were omitted from this analysis, the *F* for the Generalization effect is a direct test of the asymmetry of generalization. It was highly significant for the two experiments combined, and the *F* for the interaction of the Generalization effect with Experiments was less than unity. The significant Generalization × Tone Arrangement interaction was produced by the fact that there were

TABLE 4

ANALYSIS OF ADJUSTED NUMBER OF CRs TO NEGATIVE STIMULI IN EXP. I AND II

| Source | *df* | *MS* | *F* |
|---|---|---|---|
| Between *Ss* | 47 | | |
| Tone arrangement (T) | 1 | 1.55 | .03 |
| Hand arrangement (H) | 1 | 63.99 | 1.35 |
| Experiment (E) | 1 | 485.19 | 10.21* |
| T × H | 1 | 65.12 | 1.37 |
| T × E | 1 | .93 | .02 |
| H × E | 1 | 28.39 | .60 |
| T × H × E | 1 | 49.22 | 1.04 |
| Error (b) | 40 | 47.51 | |
| Within *Ss* | 48 | | |
| Generalization (G) | 1 | 82.10 | 15.49** |
| G × T | 1 | 53.02 | 10.00* |
| G × H | 1 | 7.10 | 1.34 |
| G × E | 1 | 3.83 | .72 |
| G × T × H | 1 | 5.19 | .98 |
| G × T × E | 1 | 15.31 | 2.89 |
| G × H × E | 1 | 3.56 | .67 |
| G × T × H × D | 1 | 3.00 | .57 |
| Error (w) | 40 | 5.30 | |
| Total | 95 | | |

* $P < .005$.
** $P < .001$.

somewhat more CRs to the high tone than to the low tone in the counterbalanced arrangement.

The assumption of the operation of both a facilitative effect and an associative, mediational effect leads to a prediction concerning the correlation across *Ss* between mean reaction time and the number of CRs for the two negative stimuli. The facilitative effect should tend to produce a negative correlation between these variables for both tones. For the Same Response stimulus, where the mediated equivalence effect is supposed to be operating, there should be a second source of negative correlation. That is, the earlier the mediating response occurs, the greater its chance of eliciting the CR within the scoring interval. This should not be the case, however, for the Different Response stimulus. This source of negative correlation should be absent, or, if a mediated distinctiveness effect is operating, there should be a source of positive correlation. This reasoning leads to the rather clear prediction that there should be a higher negative correlation between reaction time and the number of CRs for the Same Response stimulus. For Exp. I this correlation was −.24 for the Same Response tone, and +.18 for the Different Response tone. The corresponding correlations for Exp. II were −.62 and −.44, respectively. Both differences were in the predicted direction. Employing the test for significance of a difference between related correlation coefficients, neither difference was quite significant. However, when the probabilities were combined by the Fisher-Pearson chi square procedure, the difference was significant for the two experiments combined ($P = .044$, two-tailed test).

## DISCUSSION

It is the opinion of the authors that the results of these experiments should be interpreted with considerable caution. Since the facilitative effect is obviously a powerful one while the mediational one is relatively small, the danger of artifactual results must not be ignored. But in view of the important role assigned to mediated generalization in the theoretical literature, it seems important to ascertain whether it does operate at the level of the conditioned response, which is commonly regarded as a rather fundamental form of associative learning. The analyses of the conditioning measures, adjusted for their correlation with reaction time, appear to provide moderately good evidence that a mediational effect was operating in these experiments as well as the facilitative one. The verified prediction concerning the higher negative correlation between generalization and reaction time for the "equivalent" stimulus as opposed to the "distinctive" one provides corroborative evi-

dence. The acceptance of the null hypothesis concerning the interaction of the generalization effect with experiments is in accord with the prediction that the amount of mediated generalization should be independent of whether or not the generalized stimuli affect the same sensory modality as the CS.

It is obvious to suggest that mediation should be studied in the absence of facilitative effects, since statistical control of extraneous variables is never completely satisfactory. In some instances this may be possible. Yet if one wishes to study the associative consequences of a response at the time of its occurrence, one must be prepared to consider its other contemporary effects. This is not a problem specific to the present eyeblink experiments.

The strong facilitative effect which has been found in this series of experiments is of some interest and importance in its own right. Of immediate interest, is the role which it has probably played in other eyeblink studies which have involved concurrent motor responses during conditioning. For example, the study of Voeks (1954), purporting to show step-wise conditioning curves, involved a sort of two-hand reaction-time response which she conceived to be part of the true CS. The closure of two keys presented an auditory CS, the keys then being held down against a "stiff spring" for the duration of this signal. Responses were counted as CRs if they occurred anytime after the onset of the auditory stimulus. It would be difficult, indeed, to design a situation better calculated to produce a facilitative effect. This is, in part, evidenced by the fact that her "conditioning" level reached approximately 35% on Trials 2 and 3 and 60% on Trials 4 and 5. It is extremely doubtful that this situation involved conditioning to any important degree. Similar complications would be involved in such studies as those of Miller and Cole (1936) and Beck (1939).

## Summary

In two differential eyeblink conditioning experiments, *S*s made more CRs to negative stimuli accompanied by the same manual response as the CS, than they did to stimuli to which they made a different manual response. The faster the reaction time of the manual response, the more probable a CR, irrespective of which stimulus was involved. This was interpreted as evidence for a facilitation effect. The asymmetry of generalization was still present, however, when the conditioning measures were adjusted for correlation with reaction time. This was interpreted as indicating the probable action of an associative, mediated generalization effect. The correlation between number of CRs and reaction time was significantly more negative for the "equivalent" generalized stimuli than for the "distinctive" stimuli. This was also interpreted as indicating the presence of both facilitative and mediational effects.

There was no evidence that the magnitude of the mediation effect was different when generalization was from one tone to another than when it was from a light to a tone.

Implications of the facilitative effect for other eyeblink studies involving concurrent motor responses were indicated.

## References

Beck, L. H. Conditioning and the coordination of movements. *J. gen. Psychol.*, 1939, **20**, 375–397.

Grice, G. R., & Davis, J. D. Mediated stimulus equivalence and distinctiveness in human conditioning. *J. exp. Psychol.*, 1958, **55**, 565–571.

Miller, J., & Cole, L. E. The influence of a "voluntary" reaction upon the development and the extinction of the conditioned eyelid reaction. *J. genet. Psychol.*, 1936, **48**, 405–440.

Voeks, V. W. Acquisition of S-R connections: A test of Hull's and Guthrie's theories. *J. exp. Psychol.*, 1954, **47**, 137–147.

(Received June 26, 1959)

*Journal of Experimental Psychology*
Vol. 59, No. 6, 1960

# ROLES OF ASSOCIATION VALUE AND SYLLABLE FAMILIARIZATION IN VERBAL DISCRIMINATION LEARNING

WILLARD N. RUNQUIST[1] AND MADELYN FREEMAN

*Hobart and William Smith Colleges*

Paired-associate and serial verbal learning may be conceived of as consisting of two stages. The first stage consists of delimiting the population of responses, i.e., learning what the responses are. The second stage consists of pairing each response with its correct stimulus or learning the correct order of these responses. Although logically the first stage must precede the second, it is likely that the two occur concomitantly in the ordinary learning situation.

Several studies of the role of association value and familiarization of syllables in paired-associate and serial learning have suggested that factors associated with responding are more important than "perceptual" factors in acquisition of verbal lists (e.g., Cieutat, Stockwell, & Noble, 1958; Mandler & Campbell, 1957; Noble, 1952, 1955; Noble & McNeely, 1957). One hypothesis which may be advanced concerning these results is that the primary influence of association value and familiarization is in the first stage (the response-learning phase) of acquisition. According to this hypothesis, the letters making up a nonsense syllable must be integrated into a single response before correct anticipations can occur (Mandler, 1954). Syllables of high association value and familiarity are conceived of as representing verbal habits established outside the experiment and thus are more readily integrated. The number of responses available on any given trial is then determined by how far this integration stage has progressed.

However, suppose a learning task is used in which the total number of available responses is determined by the task, rather than the rate of response integration. This may be accomplished by using a procedure called verbal discrimination learning, where *S* is presented with a pair of syllables and must select the response arbitrarily designated as "correct" from these two. The response population from which *S* must make his choice is thus restricted to the two syllables presented, and integration in the sense described above is unnecessary. It might then be expected that neither association value nor familiarization would have much effect upon acquisition in this task.

The following experiments are designed to investigate the role of both association value and familiarization on the verbal discrimination learning of nonsense syllables.

## EXPERIMENT I

The first two experiments were concerned with the effects of association value. Two methods of varying association value are available for verbal discrimination learning: A pair may be made up of one high and one low association value syllable, or the two syllables making up a pair may both be high or both low association value. According to the hypothesis, there seemed to be no reason to favor one method or the other. Although this conclusion later appeared to be

[1] Now at Pomona College.

in error, for the first experiment the first method (heterogeneous pairs) was used.

### *Method*

*Design and materials.*—Two sets of eight pairs of syllables were constructed from Glaze's lists (1928). In each pair one syllable was 80% to 100% association value and one syllable was 0% to 27% association value. Formal similarity was kept to a minimum both within and between sets, and the same number of letters was duplicated within a set. These sets were then combined to make a list of 16 pairs of syllables. For half of the *Ss*, the high association value syllables of Set I and the low association syllables of Set II were correct. For the other half of the *Ss*, the low association value syllables of Set I and the high association value syllables of Set II were correct. This reduced the chances of extraneous factors such as differential difficulty or interference with specific syllables influencing the results. It was felt that having both high and low association value syllables correct in the same list would avoid *Ss* recognizing, for example, that all the "word-like" ones were correct.

*Procedure.*—The 16 pairs in each list were arranged in three orders and presented with a 16-mm. filmstrip projector. The pairs were flashed from the back of the screen so that *S* could not see *E* or the projector during the presentation. General illumination was reduced to minimized distraction. Syllables appeared black on a white background and were ¾ in. high. They appeared on an 8 × 8-in. milk glass piece in the center of a 3 × 3-ft. wooden screen. The two syllables were presented one above the other for 2 sec. followed by the correct syllable for 2 sec. The intertrial interval was 4 sec. The position of the syllables in a pair varied from trial to trial. All *Ss* were given 15 trials unless they went three consecutive trials with no error. They were then stopped and given a perfect score for the remaining trials. The *Ss*, 20 volunteers from Hobart and William Smith Colleges, were instructed to select and spell aloud the syllable designated as correct on each presentation including the first trial. All *Ss* were naive to verbal learning experiments.

### *Results*

The percentages of correct anticipations for the high association value syllables and low association value syllables on each trial are shown in Fig. 1. The hypothesis that variations in association value would not produce performance differences is not supported by these data. Statistically, the mean total correct anticipations for the high association value syllables is significantly greater than for the low association value syllables ($t = 3.65$). Seventeen of the 20 *Ss* gave more correct responses to the high than to the low association syllables.

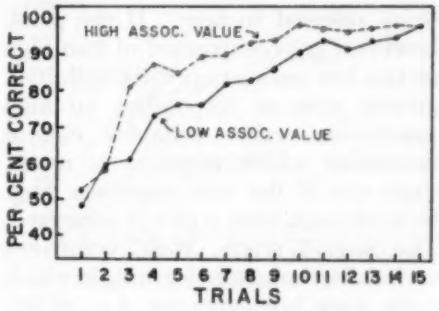



FIG. 1. Percentage of correct responses on each trial for high and low association value syllables in Exp. I (heterogeneous pairs).

## EXPERIMENT II

Rather than completely reject the hypothesis on the basis of this single experiment, explanation was sought in terms of other factors associated with making the response. The facilitation in performance obtained here may have been due to greater ease of responding with the high association syllables. Essentially this notion is what Noble (1955) has called "motor patterning." If it is easier to learn to make the high association value responses than the low association value responses, some facilitation might be expected. This is a kind of response integration although only the motor aspect is

being referred to here. If the pairs, however, are constructed of two high or two low association value syllables, greater ease of responding to high association value syllables cannot determine which response is made, since one of the two responses must be made each time a pair is presented. The second study, then, compared verbal discrimination learning in which pairs were homogeneous, i.e., either both syllables had high or both syllables had low association value.

*Method*

The same syllables were used as in Exp. I, except that they were re-paired to form two sets of eight syllables in which Set I contained only high association value syllables and Set II only low association value syllables. Duplicated letters were again kept at a minimum and equalized within each set. Two lists were used consisting of the two sets combined, with the correct member of each pair in List 1 being the incorrect member of the same pair in List 2. Half of the *Ss* learned each list.

The *Ss* were 20 volunteers from Hobart and William Smith Colleges. All were naive with respect to verbal learning. The procedure was identical to that in Exp. I.

*Results*

The acquisition curves for the high and low association value pairs are presented in Fig. 2. Although there

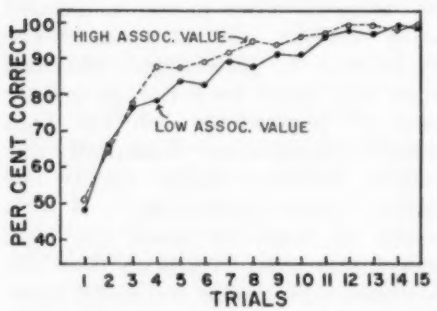



FIG. 2. Percentage of correct responses on each trial for high and low association value syllables in Exp. II (homogeneous pairs).

is some evidence for better performance on the high association value pairs, the differences are much smaller than those obtained in Exp. I. Comparison of the mean total correct responses for the two kinds of syllables resulted in a $t$ of 2.07, which is just short of significance at the .05 level for 19 *df*. Although a constant number of trials was used, all *Ss* obtained all eight items within each set correct on at least one trial. Thus trials to 8/8 criterion may be used as a measure of acquisition rate. Seventeen of the 20 *Ss* reached this criterion first on the high association value pairs. A comparison of the mean trials to this criterion for the two kinds of pairs resulted in a $t$ of 3.45. It thus appears that the difference in performance shown in Fig. 2 is reliable, although a small one.

A comparison was made between Exp. I and II of the mean difference between the total correct responses to high and low association value pairs. The value of $t$ obtained was 1.73 which is only significant at about the .08 level.

The results of Exp. I and II do not support the hypothesis that the facilitative effects of association value can be attributable solely to the response-learning phase. Although there are several interpretations of these differences, the one favored here is suggested by Underwood and Richardson (1958). High association value syllables may provide some sort of extra cue value, often called "meaning," thus making them less susceptible to interference from duplicated letters. Since in the lists used in these experiments it was impossible to avoid some duplication of letters, it is not unreasonable to assume that the performance on the low association items in Exp. II was depressed by the slight amount of interference present. Some support for this position comes from the fact that a pilot

study using 10 homogeneous pairs in which there were no duplicated letters, except a few vowels, produced no differences between high and low association value pairs ($t = 1.27$ with 19 $df$). The present experiments used lists of 16 pairs because it was felt that acquisition was too rapid with a 10-item list to give any differences a chance to appear. Increasing the length of the list necessarily increases the duplicated letters due to the restricted number of letters and syllables available, thus increasing interference. The difference obtained in Exp. II, while not significantly smaller than that obtained in Exp. I, tends to support the hypothesis that the difference in the latter experiement was mostly due to differential ease of making the response.

## Experiments III and IV

The purpose of these two studies was to compare the effects produced by the familiarization of nonsense syllables (Noble, 1955) with those produced by association value.

### *Method*

*Design and materials.*—The rules for the construction of lists were the same as those in the previous experiments, with the exception that all syllables were from 0% to 27% association value (Glaze, 1928). The number of duplicated letters within a set was slightly higher for these lists than in Exp. I and II. As in the previous studies, a given syllable in each pair was correct for half the *S*s and incorrect for the other half.

*Procedure.*—Prior to learning the list each *S* was given familiarization training on half the syllables.

Half the syllables were correct in learning and half the syllables were incorrect For Exp. III, one syllable was chosen from each pair; for Exp. IV, both syllables of half of the pairs were used. Two different familiarization lists were employed in each experiment so that syllable difficulty, etc., was partially counterbalanced between the familiarized and nonfamiliarized syllables. In familiarization training each syllable was exposed singly on the screen for 2 sec. The *S* was instructed to spell out each syllable as it appeared. The

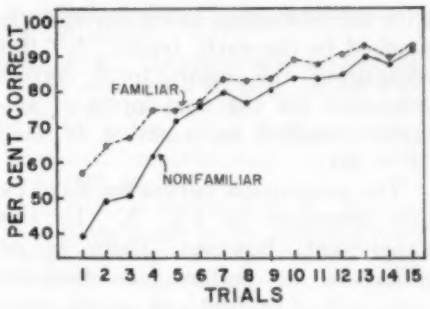



FIG. 3. Percentage of correct responses on each trial for familiar and nonfamiliar syllables in Exp. III (heterogeneous pairs).

16 syllables were arranged in six different orders, but there was no interval between orders. No syllable preceded or followed any other syllable more than once. Each syllable was seen and spelled 12 times by each *S*.

The verbal discrimination learning procedure was identical with that in Exp. I and II. Twenty volunteers from the Hobart and William Smith students served as *S*s in each study. None of the *S*s had served in a verbal learning experiment before.

### *Results*

The percentage of correct responses on each trial for familiarized and nonfamiliarized syllables in Exp. III is shown in Fig. 3. With heterogeneous pairs performance level is higher when the familiarized syllable is correct than when it is incorrect,

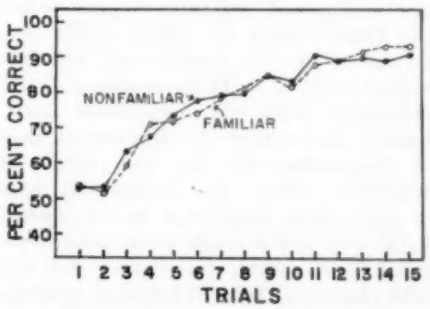



FIG. 4 Percentage of correct responses on each trial for familiar and nonfamiliar syllables in Exp. IV (homogeneous pairs).

with the advantage being particularly marked in the early trials. A $t$ test comparing the mean total correct responses for the two kinds of syllables resulted in a value of 3.78 ($P < .01$).

The acquisition curves for Exp. IV are presented in Fig. 4. In this experiment, however, there is no difference in performance between pairs of familiarized and nonfamiliarized syllables. Comparison of the mean total correct responses gave a $t$ of .33, which is clearly not significant.

## Discussion

The effects of familiarization training on verbal discrimination learning are not the same as the effects of association value. The fact that familiarization produced facilitation in the case of heterogeneous pairs is consistent with the effects produced by association value, although there are some slight differences. In the case of association value, performance on the two kinds of pairs starts at the same level, then diverges before converging again near the asymptote. Familiarization, on the other hand, produces large differences on the first trial. This large initial difference in the case of familiarization is probably produced by a tendency to guess and to keep guessing the syllables which have just been seen. This is clearly shown by the fact that performance on the nonfamiliarized syllables begins below the chance level (at 39%), while performance on the familiarized syllables begins at 57%. These results are also consistent with the hypothesis attributing facilitation to differential ease of responding for the two kinds of syllables. Since the differential ease of responding might not be established until after a few trials when association value was varied, facilitation did not take place until later in learning, whereas familiarization training established these habits prior to learning the list.

The striking result of these studies, however, is the failure of familiarization training to produce differences in performance when homogeneous pairs are used. This result is also consistent with the interpretation given the association value experiments. If the differences which occurred in the homogeneous pairs association value experiment are attributed to the cue value of association or "meaning," familiarization training should not produce differences since it can be shown that familiarizing operations, in general, do not increase association value (Riley & Phillips, 1957). It might be argued that a difference would have been obtained in Exp. IV had familiarization training been continued. This cannot, of course, be denied; however, the 12 trials given here were sufficient to produce large differences in Exp. III when conditions for producing these differences were allowed. In fact, the purpose of Exp. III was to ensure that the familiarizing operations were sufficient to produce performance differences in one situation where these differences might be expected.

It is felt that, over-all, the results support a response-oriented interpretation of the effects of association value and familiarization, although they do indicate some cue effect of association value. The results of Exp. IV represent the only real support for the original predictions from the response integration hypothesis that association value and familiarization would not facilitate verbal discrimination learning. Nevertheless, an attempt has been made to show that complicating factors were present in the case of heterogeneous pairs.

## Summary

Four verbal discrimination learning experiments were performed to test the hypotheses that association value and familiarization with the syllables would not facilitate performance. Two experiments varied association value and two gave familiarization training on the syllables. In one experiment each pair of items to be learned consisted of one high and one low association value syllable (heterogeneous pairs), while in the other experiment, both syllables were either high

or low association value. The same held true for the familiarization studies.

The results showed that both association value and familiarization facilitated performance with the heterogeneous pairs, while only association value facilitated performance with the homogeneous pairs. These results were discussed in relation to an interpretation of the role of association value and familiarization in learning in terms of their effects on response availability

## References


CIEUTAT, V. J., STOCKWELL, F. E., & NOBLE, C. E. The interaction of ability and amount of practice with stimulus and response meaningfulness ($m$, $m'$) in paired-associate learning. *J. exp. Psychol.*, 1958, **56**, 193–202.

MANDLER, G. Response factors in human learning. *Psychol. Rev.*, 1954, **61**, 235–244.

MANDLER, G. H., & CAMPBELL, E. H. Effects of variation in associative frequency of stimulus and response members on paired-associate learning. *J. exp. Psychol.*, 1957, **54**, 269–273.

NOBLE, C. E. The role of stimulus meaning ($m$) in serial verbal learning. *J. exp. Psychol.*, 1952, **43**, 437–446.

NOBLE, C. E. The effect of familiarization upon serial verbal learning. *J. exp. Psychol.*, 1955, **49**, 333–338.

NOBLE, C. E., & MCNEELY, D. A. The role of meaningfulness ($m$) in paired-associate verbal learning. *J. exp. Psychol.*, 1957, **53**, 16–22.

RILEY, D. A., & PHILLIPS, L. W. The influence of syllable familiarization on association value, serial rote learning, and reminiscence. *Amer. Psychol.*, 1957, **12**, 425. (Abstract)

UNDERWOOD, B. J., & RICHARDSON, J. Studies of distributed practice: XVIII. The influence of meaningfulness and intralist similarity of serial nonsense lists. *J. exp. Psychol.*, 1958, **56**, 213–219.


(Received June 26, 1959)

*Journal of Experimental Psychology*
Vol. 59, No. 6, 1960

# EFFECT OF RUNWAY SIZE AND DRIVE STRENGTH ON ACQUISITION AND EXTINCTION[1]

DONALD J. LEWIS

*Louisiana State University*

AND JOHN W. COTTON

*Northwestern University*

The theoretical importance of competing responses has been repeatedly emphasized in the past (Guthrie, 1935; Wendt, 1936), and data are beginning to accumulate indicating that competing responses do affect ordinarily measured behaviors in important ways. Cotton (1953) showed that a major part of the effect of drive on asymptotic running speed was due to the intervention of competing responses. When trials on which competing responses occurred were not analyzed, drive did not have as great an effect on performance as it did when competing response trials were included in the analysis. Pereboom and Crawford (1958) showed essentially the same effect for magnitude of reward and number of acquisition trials.

Adelman and Maatsch (1955, 1956) have shown that controlling the competing responses during extinction results in a marked effect upon behavior. When they allowed *S*s to climb out of the goalbox, running speeds were increased during extinction. This finding was confirmed by Cotton, Lewis, and Metzger (1958) and extended to acquisition behavior for both the runway and the T maze.

Sheffield (1950), Stanley (1952), Amsel (1958), and others have emphasized the drive-arousing properties of the extinction operation, and others (Spence, 1956; Taylor, 1951) have indicated the connection between drive and competing responses.

There seems little doubt, then, about the importance of competing responses. And the effect of competing responses in conjunction with drive is also of interest. The present paper is a report of an experiment in which two conditions of drive and two conditions of competing responses were manipulated factorially. The drive conditions were established by hours of food deprivation; the competing response conditions were established by two runway sizes, large (L) and small (S). It was assumed that the L condition would permit many competing responses, whereas the S condition would tend to prevent the competing responses from occurring. In the L runway *S* could turn around, scratch himself, move from one side of the alley to the other, and perform many responses other than running to the goalbox. The S runway, it was believed, would prevent most of these competing responses.

## METHOD

*Subjects.*—The *S*s were 80 naive male albino rats of the Sprague-Dawley strain. They varied from 66 to 81 days of age, with a mean of 74.3 days. Because it was necessary to make structural adjustments in the apparatus for different experimental groups, departures from purely random assignment of *S*s were necessary. When *S*s arrived, they

[1] This study was supported by a grant from the National Science Foundation.

The writers wish to express their appreciation for the work of Glen Jensen who participated in all phases of the present investigation.

were assigned by a table of random numbers to their individual living cages. Then they were assigned to the experimental conditions in blocks of four with each block of four serving under identical conditions. Four *Ss* were run in each of the four experimental conditions before any one of the experimental conditions received its second assignment of four *Ss*. Replications continued in this fashion until 20 *Ss* were assigned to each of the experimental conditions. Four *Ss* were discarded because of a procedural error, and they were replaced.

*Apparatus.*—The apparatus was a straight runway, the inside dimensions of which could be altered by putting in or taking out Plexiglas inserts. Two adjustments of the inside dimensions were employed. In one case the runway was set at its maximum width and depth. In this L condition the inside dimensions of the runway proper were 8½ in. high, 4 in. wide, and 4 ft. long. The inside dimensions of the goalbox were 8½ in., 9 in., and 10½ in. The goalbox also contained a recess, 2¾ in. by 1¾ in., in which the foodcup was placed so that it could not be seen before a complete goalbox entry. For the startbox the dimensions were 8½ in., 4 in., and 6 in. The inside dimensions of the startbox for the S condition were 2½ in., 1⅝ in., and 6 in. For the runway proper the dimensions were 2½ in., 1⅝ in., and 4 ft. The goalbox was 2½ in. high but remained unchanged in the other dimensions. The interior of the apparatus was painted flat black. Response time was defined as the interval required from the opening of the startbox door until *S* broke a photoelectric beam 2 in. from the goalbox. The distance between the startbox and the photoelectric beam was 44½ in. Times were taken automatically by a Standard Electric timer, calibrated in .01 sec.

*Design.*—Two variables were manipulated factorially. One was the number of hours of food deprivation and the other was the inside dimensions of the apparatus. One-half of the *Ss* were assigned to the L runway for acquisition and extinction, and one-half were assigned to the S runway. One-half of the *Ss* run in the L apparatus were deprived of food for 2–3 hr. at the beginning of any experimental session. The remaining half of the *Ss* were deprived approximately 22 to 23 hr. The same deprivation conditions were applied to *Ss* run in the S apparatus.

*Deprivation and taming.*—All *Ss* were placed on their appropriate deprivation schedule at least seven days before the initiation of experimentation. The *Ss* were handled for at least 5 min. on each of these seven days and made familiar with the 45-mg. lab chow pellets to be used during the acquisition phase of the experiment.

*Acquisition.*—Each *S* received 10 trials a day for five days. Then two days intervened during which *S* received no trials, but was maintained on the appropriate deprivation schedule. There followed another 10 trials a day for five days. A total of 100 acquisition trials was given with approximately a 15-sec. intertrial interval and with a 45-mg. pellet available in the goalbox on each trial. The *Ss* were detained in the goalbox approximately 10 sec. on each trial.

*Extinction.*—Extinction began two days after the one-hundredth acquisition trial. During these two days *Ss* were maintained on their appropriate deprivation schedule. Ten nonreinforced trials a day for five successive days were given during the extinction period. If on one of these trials *S* did not reach the goalbox within 1 min., he was removed from the apparatus and the response was recorded as 1 min. The intertrial interval remained at approximately 15 sec., and *Ss* continued to be detained in the goalbox for approximately 10 sec. on each trial.

## Results

*Acquisition.*—The reciprocal of the running time (speed) for each trial was determined for each *S*, and all analyses were performed on these speeds. A graph of the mean speed by 10-trial blocks is presented in Fig. 1. An over-all analysis of variance was performed on the means

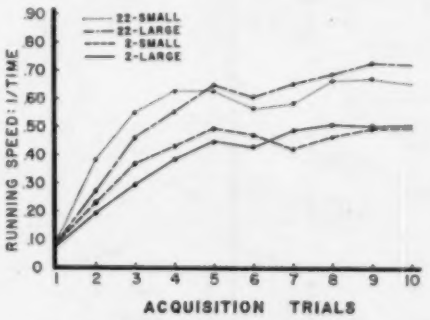



FIG. 1. The relationship between acquisition trials and running speed. Each of the 10 points on the abscissa represents the mean of 10 trials. Speeds are the reciprocals of running time in seconds.

of the 100 acquisition trials. The apparatus size was found to be *not* an effective variable ($F = .32$, $df = 1/76$). Drive was significant ($F = 15.18$, $df = 1/76$); and there was no interaction between apparatus size and drive ($F = .03$, $df = 1/76$). Figure 1 shows the clear superiority of the high-drive groups throughout acquisition. It also seems clear from Fig. 1 that the two drive conditions resulted in different asymptotes. The slight dip in all four acquisition curves just after Trial 50 can probably be explained by the two-day interval after Trial 50 during which *S*s received no training.

The over-all analysis of variance, which shows apparatus size not to be an effective variable, is probably not completely appropriate for these data because the analysis includes values for all trials in the computations of means for each group. The initial elevation of the S groups is counterweighted by the later elevation of the L groups, and the over-all analysis is insensitive to this change as a function of trials. Therefore, another analysis of variance was performed with trials as an added variable. The means of 10-trial blocks 31–40, 41–50, 51–60, and 61–70 were taken for each *S* and these scores made up the added variable of trials. This analysis confirms what Fig. 1 seems to show: that the effect of apparatus size is a function of the number of acquisition trials. The *F* for the interaction between apparatus size and trials was 5.53. A value of 3.97 is required for significance at the .01 level with 3 and 114 *df* (the pooled *S*s × Trials served as an error term).

Perhaps, because of its *post hoc* nature, this "repeated measurements" analysis capitalized on observed chance variation in the data. The alternative is to be blinded by a poor first guess as to the appropriate analyses. The final test of any empirical finding is its reliability over several experiments.

A record was kept of all the trials on which competing responses occurred. A competing response was defined as any behavior which did not result in forward progress through

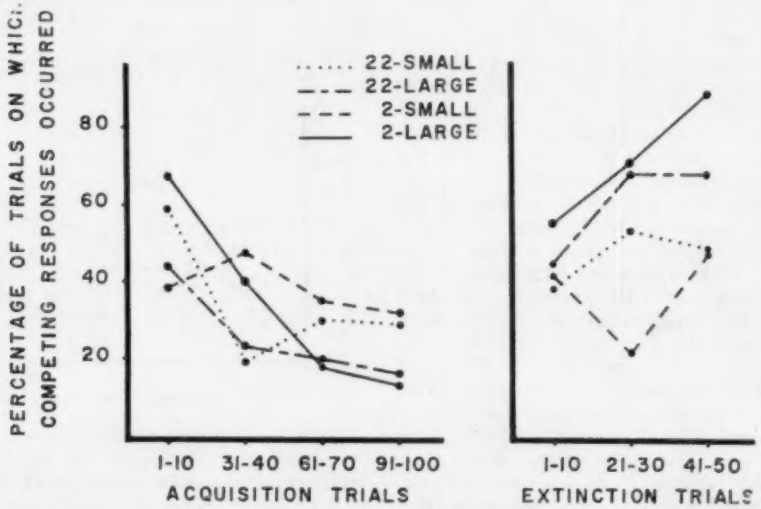



FIG. 2. The percentage of trials on which one or more competing responses occurred.

TABLE 1

ANALYSIS OF THE CHANGE IN COMPETING RESPONSES WITH TRIALS DURING ACQUISITION AND EXTINCTION

| Source | *df* | Acquisition | | Extinction | |
|---|---|---|---|---|---|
| | | *MS* | *F* | *MS* | *F* |
| Drive (D) | 1 | 46 44 | 2.07 | 18 | — |
| Apparatus size (A) | 1 | 15 | — | 217 | 11.88* |
| D × A | 1 | 7 | — | 7 | — |
| Between *Ss* in the same group | 76 | 22.25 | | 18.4 | |
| Trials (T) | 3 | 141.6 | 33.2* | 108.5 | 30.74* |
| T × D | 3 | 21.3 | 4.99* | 1.56 | — |
| T × A | 3 | 18.3 | 4.28* | 27 | 7.65* |
| T × D × A | 3 | 36.3 | 8.50* | 4 | 1.13 |
| Pooled *Ss* × T | 228 | 4.3 | | 3.53 | |

* $P = .01$.

the maze. Such behaviors as retracing, face washing, and scratching were ruled as competing responses. Figure 3 shows the percentage of trials on which one or more competing responses occurred as a function of the manipulated variables, and Table 1 shows the result of "repeated measures" analysis of these data.

During acquisition none of the variables resulted in significant differences by themselves, but they both interacted significantly with trials. In addition, there was a significant triple interaction of the two variables with trials. If the mean square for the triple interaction is used as the error terms in testing the other differences, only the *F* for Trials remains significant. The Trials *F* means that the number of competing responses decreases as acquisition proceeds.

The triple interaction seems to be attributable to the differential effect over trials of low drive, depending on the apparatus that is used. In the L apparatus the low-drive group yields a higher percentage of trials with competing responses early in acquisition than does the high-drive group. Late in acquisition, however, the low-drive group in the large apparatus has very few competing responses, whereas the low-drive group in the small apparatus has many.

To state the meaning of this triple interaction in another way: Apparatus has no differential effect over trials for the high-drive group. And, in general, fewer competing responses occur in the S apparatus throughout acquisition. The effect of drive, however, is reversed from early to late acquisition for the L apparatus groups. Early in acquisition more competing responses occur in the L apparatus, and late in acquisition more competing responses occur in the S apparatus. Decreasing apparatus size decreased the percentage of trials on which competing responses occurred only very early in acquisition for the low-drive group. Other than that, more competing responses occurred in the S apparatus.

*Extinction.*—An over-all analysis of variance of the extinction running speeds was performed. Both Drive ($F = 19.1$ $df = 1/76$) and the Apparatus ($F = 13.5$, $df = 1/76$) were highly significant variables. There was no interaction between the two ($F = .05$, $df = 1/76$). The interpretation of this analysis can be obtained from Fig. 3. This figure indicates that the S apparatus resulted in *increased* resistance to extinction, and that the greater drive also resulted in prolonged extinction. An analysis of covariance in which the mean of the last 10 acquisition trials was used as the control variable and the mean of all extinction trials as the prediction variable gave approximately the same results. The *F* for the apparatus was increased to 28.86 for 1 and 75 *df*, however. An analysis of covariance, again with the mean of the last 10 acquisition trials as the control variable, but with the mean of the last 10 extinc-

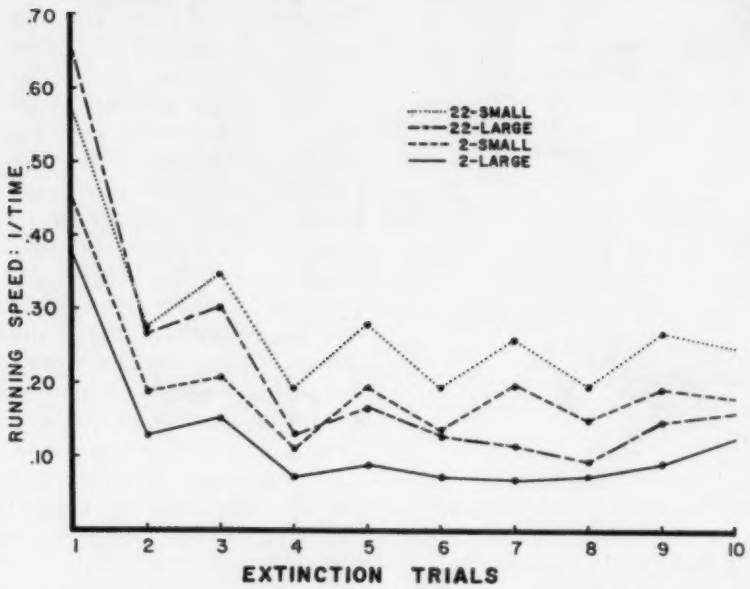



FIG. 3. The relationship between extinction trials and running speed. Points 1 to 2, 3 to 4, 5 to 6, 7 to 8, and 9 to 10 represent blocks of 10 extinction trials given on each of five days. Points 2 to 3, 4 to 5, 6 to 7, and 8 to 9, represent spontaneous recovery between 10-trial extinction blocks.

tion trials as the predicted variable, shows the Apparatus as remaining significant ($F = 10.10$, $df = 1/75$).

Finally, a repeated measurements analysis was performed using the mean of the last 10 acquisition trials, the mean of extinction Trials 21 to 25, which form the midpoint of the extinction trials, and the mean of the last 5 extinction trials as points for the trend. A summary of this analysis is presented in Table 2. Drive alone remains significant; Apparatus is not significant. Because all the other extinction analyses show Apparatus to be a highly significant variable, this low $F$ value is perhaps surprising. The reasons for the nonsignificant $F$ for Apparatus can be seen by looking at Fig. 1 and 3 jointly. At the end of acquisition, shown in Fig. 1, the mean of the S conditions is below that of the L conditions. During extinction, shown in Fig. 3, just the reverse is true; the mean of the S conditions is above that of the L conditions. This reversal cancels out the over-all effect of the apparatus. The reversal also means that there should be a significant interaction of apparatus with trials. The very significant $F$

TABLE 2

ANALYSIS OF THE CHANGE IN EXTINCTION RUNNING SPEEDS WITH TRIALS

| Source | *df* | *MS* | *F* |
|---|---|---|---|
| Drive (D) | 1 | .741 | 17.50* |
| | 1 | .157 | 3.68 |
| Apparatus size (A) | 1 | .001 | .03 |
| D × A | 76 | .043 | |
| Between *Ss* in the same group | | | |
| Trials (T) | 2 | 4.556 | 236.32* |
| T × D | 2 | .100 | 5.20* |
| T × A | 2 | .122 | 6.37* |
| T × D × A | 2 | .005 | .28 |
| Pooled *Ss* × T | 152 | .019 | |

* $P = .01$.

for the Trials × Apparatus interaction confirms this interpretation. The S apparatus resulted in slower running speeds at the end of acquisition but in faster running speeds throughout most of extinction.

The interaction between drive and trials is also significant. This significance is attributable to the general decrease in the effectiveness of drive as extinction progresses.

Figure 3 and an analysis of variance show that during extinction more competing responses occurred in the L apparatus, and that drive had no effect on competing responses.

*Spontaneous recovery.*—Ten extinction trials were run on the first extinction day, and 10 trials a day were given for four succeeding days. To determine the amount of spontaneous recovery, the *difference* between the speeds on the tenth extinction trial and the eleventh extinction trial was computed, and an analysis of variance performed on these difference scores. Drive was very significantly effective ($F = 10.4$, $df = 1/76$), Apparatus size was not ($F = 1.2$), and there was no significant interaction between the two ($F = .14$). Figure 3 shows clearly that the higher the drive, the more the spontaneous recovery.

## Discussion

As in a study by Cotton, Lewis, and Metzger (1958), a wide runway was accompanied by an increased number of trials on which competing responses occurred. Because more competing responses usually meant slower running times, rapid extinction corresponded with an increase in competing responses. The two measures are, however, to some extent independent. The *S*, for example, can run fast or slow on trials on which no competing responses occur. Also, it is possible for a competing response to occur on a fast trial as well as on a slow trial. Therefore, it is meaningful to conclude that competing responses are at least one determiner of extinction, as measured by decreased running times, and this conclusion is in line with a number of interference theories of extinction.

That the S apparatus, in general, increased the number of competing responses was a surprise, although there was at least a suggestion of such a result in the previous Cotton, Lewis, and Metzger study. Evidently apparatus size has a different effect on competing responses during acquisition than during extinction. And it is clear that a small apparatus, little bigger than a rat's body, does not necessarily curtail competing responses. The initial plan to manipulate competing responses by restricting the area for locomotion was a failure.

## Summary

Two conditions of drive—2 and 22 hr. of food deprivation—were combined factorially with a large and a small runway in an attempt to determine some of the effects of drive on competing responses. The length of the runway was the same in both apparatus conditions, but the small runway was constricted greatly in other dimensions in an attempt to reduce competing responses. Twenty animals were assigned to each of the four conditions. They were given 100 acquisition and 50 extinction trials. The *S*s were run under the same experimental conditions during extinction that they had experienced during acquisition.

The major findings were:

1. High-drive resulted in faster running speeds throughout acquisition and early extinction.

2. Early in acquisition the small apparatus resulted in faster speeds; there is a reversal of this trend later in acquisition.

3. The small apparatus decreased the number of trials on which competing responses occurred only for the first few trials of acquisition and then only for the low-drive groups. Other than this, more competing responses occurred in the small apparatus during acquisition.

4. During most of extinction the small apparatus produced greater running speeds than the large, although this superiority was lost by the end of extinction.

5. The high-drive condition resulted in greater spontaneous recovery.

6. In extinction, the small apparatus served to reduce competing responses.

## REFERENCES

ADELMAN, H. M., & MAATSCH, J. L. Resistance to extinction as a function of the type of response elicited by frustration. *J. exp. Psychol.*, 1955, **50**, 61–65.

ADELMAN, H. M., & MAATSCH, J. L. Learning and extinction based upon frustration, food reward, and exploratory tendency. *J. exp. Psychol.*, 1956, **52**, 311–315.

AMSEL, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychol. Bull.*, 1958, **55**, 102–119.

COTTON, J. W. Running time as a function of amount of food deprivation. *J. exp. Psychol.*, 1953, **46**, 188–198.

COTTON, J. W., LEWIS, D. J., & METZGER, R. Running behavior as a function of apparatus and of restriction of goal box activity. *J. comp. physiol. Psychol.*, 1958, **51**, 336–341.

GUTHRIE, E. R. *The psychology of learning.* New York: Harper, 1935

PEREBOOM, A. C., & CRAWFORD, B. M. Instrumental and competing behavior as a function of trials and reward magnitude. *J. exp. Psychol.*, 1958, **56**, 82–85.

SHEFFIELD, V. F. Resistance to extinction as a function of distribution of extinction trials. *J. exp. Psychol.*, 1950, **40**, 303–313.

SPENCE, K. W. Behavior theory and conditioning. New Haven: Yale Univer. Press, 1956.

STANLEY, W. C. Extinction as a function of the spacing of extinction trials. *J. exp. Psychol.*, 1952, **43**, 249–260.

WENDT, G. R. An interpretation of inhibition of conditional reflexes as competition between reaction systems. *Psychol. Rev.*, 1936, **43**, 258–81.

(Received June 26, 1959)

*Journal of Experimental Psychology*
Vol. 59, No. 6, 1960

# LEARNING AND RETENTION OF PAIRED ADJECTIVES AS A FUNCTION OF PERCENTAGE OCCURRENCE OF RESPONSE MEMBERS

RUDOLPH W. SCHULZ [1] AND WILLARD N. RUNQUIST [2]

*Northwestern University*

If the paired-associate verbal learning situation is considered an analogue of the classical conditioning situation, then the presentation of the response member of the S-R pairs corresponds to the presentation of the UCS and reinforcement. Therefore variation in percentage occurrence of response members in the paired-associate situation should be analogous to variation in the percentage of reinforcement in the conditioning situation.

At the time the present study was begun, percentage occurrence of response members (hereafter abbreviated ORM) had not been investigated in paired-associate verbal learning. Recently, Goss, Morgan, and Golin (1959) have reported that the mean number of trials to learn paired nonsense-syllable lists to criterion decreased as percentage ORM increased. Adjective lists were used in the present experiment to investigate the ORM dimension at five points; namely, 20%, 40%, 60%, 80%, and 100% ORM.

It is well established that intermittently reinforced responses are typically more resistant to extinction than continuously reinforced responses. If an unlearning (Melton & Irwin, 1940; Melton & von Lackum, 1941) extinction-like (Underwood, 1948) process underlies forgetting, at least in part, then it is conceivable that a paired-associate list learned with intermittent ORM will be more resistant to forgetting than a list learned with continuous ORM. Therefore a second purpose of the present experiment was to study retention as a function of percentage ORM.

[1] Now at Carnegie Institute of Technology, Pittsburgh, Pennsylvania.

[2] Now at Pomona College, Claremont, California.

## METHOD

*List construction.*—All *Ss* in all conditions learned the same 10 pairs of adjectives. Intralist similarity among these pairs was low. Details concerning the original selection of the pairs may be found elsewhere (Underwood, 1951). Five lists, differing only with respect to percentage ORM, were constructed utilizing these 10 pairs of adjectives. The method of manipulating percentage ORM was essentially identical with that used by Goss, Morgan, and Golin (1959); hence there is no need to detail it here. In short, the occurrence of the response members of the adjective pairs was scheduled so that percentage occurrence per trial for a given list was the same as the percentage occurrence for a given pair in the list within successive blocks of five trials. For example, with 20% ORM, 2 of the 10 stimuli were accompanied by their response members on each trial. The remaining eight stimuli were accompanied by a blank. On one of the five trials within each successive block of five trials each stimulus was accompanied by its response member once or 20% of the time. While Goss et al., used an 8-item list of nonsense syllables with 25%, 50%, 75%, and 100% ORM, the present study employed a 10-item list of adjectives with 20%, 40%, 60%, 80%, and 100% ORM.

*Subjects and procedure.*—The 20%, 40%, 60%, 80%, and 100% ORM lists were learned by five different groups of 50 *Ss* each. The 250 *Ss* were students at Northwestern University taking introductory psychology courses. They were naive to the learning of paired adjectives, although some had served in other verbal learning experiments. The *Ss* were assigned to lists in rotation as they appeared at the laboratory.

The *Ss* learned by the anticipation method. All *Ss* were given standard paired-associate learning instructions. In addition, *Ss* learning the intermittent ORM lists were warned that the response members of the pairs would be omitted a certain proportion of the time and that they should do their best in spite of this handicap.

The lists were presented on a Patterson memory drum at a 2:2-sec. rate. The stimulus term was presented for 2 sec. while a shutter covered the response term. Then the shutter lifted and both the stimulus and the response (or a stimulus and a blank on non-ORM trials) remained in the aperture for an additional 2 sec. The intertrial interval was 4 sec. The criterion for learning and relearning was one perfect recitation of the list. All relearning was with the 100% ORM list. The *Ss* reported to the laboratory at the same time on two successive days. Acquisition took place on the first day and relearning on the second day. Hence the average interval between the end of acquisition and the beginning of relearning was approximately 23.5 hr.

A total of 24 *Ss* had to be eliminated from the experiment for various reasons. Elimination rate was not differential with respect to conditions, except in the case of failure to understand the instructions for learning. The seven *Ss* dropped for this reason were all in either the 20%, 40%, or 60% ORM groups. Therefore, if ability to understand instructions is highly correlated with ability to learn paired adjectives, then the sample of *Ss* in these groups may be slightly superior to the *Ss* in the 80% and 100% ORM groups.

*Statistical analysis.*—Analysis of variance was used when the scores under consideration met the requisite assumptions. The *F*s have 4 and 245 *df* in each comparison, with an *F* of 2.37 required for significance at the 5% level. If either heterogeneity of variances or indications of nonnormality were present, then the median test (Dixon & Massey, 1957, p. 296) was used. The $\chi^2$ in this test has 4 *df* in each comparison, and a $\chi^2$ of 9.49 is required for significance at the 5% level.

## Results

*Acquisition.*—The present results, for measures identical to those reported by Goss et al. (1959), are shown in Table 1.[3] Their results for 8-item syllable lists learned to a $\frac{7}{8}$ criterion at a 2:2-sec. rate of presentation are also presented in Table 1 to facilitate the comparison of the two studies. For the present study, it may be seen from Table 1 that the mean number of trials and correct responses in learning to criterion generally decreased as percentage ORM increased. The "deviant" performance of the 60% ORM group is the only exception to an otherwise monotonic relationship for both measures. The differences among these means are statistically reliable (for trials $F = 4.07$, $df = 4/245$, $P < .01$; for correct responses $F = 2.68$, $P < .05$). It is also apparent from Table 1 that the present results agree well with those obtained by Goss et al., though the differences between the means in the

TABLE 1

Comparison of Goss, Morgan, and Golin Results with Schulz and Runquist Results

| Study | Percentage ORM | Means | |
|---|---|---|---|
| | | Trials | Correct Responses |
| Goss, Morgan, & Golin ($N = 20$) | 100 | 24.8 | 59.4 |
| | 75 | 34.3 | 70.2 |
| | 50 | 40.6 | 94.0 |
| | 25 | 40.5 | 102.0 |
| Schulz & Runquist ($N = 50$) | 100 | 17.5* | 93.2 |
| | 80 | 18.7 | 102.5 |
| | 60 | 16.8 | 91.0 |
| | 40 | 21.8 | 104.5 |
| | 20 | 22.7 | 113.7 |

* $\sigma_m$ based on within groups variance estimate was, for trials = 1.22, and for correct responses = 5.62. Corresponding *F*s were 4.07 ($P < .01$) and 2.68 ($P < .05$).

[3] A third measure, mean reinforced correct responses, was reported graphically by Goss et al., (1959). Since these means are readily estimated from the data in Table 1 by multiplying the mean number of correct responses for a given group by the proportion of ORM it received, they are not reported here. In accord with Goss' finding, the present mean number of reinforced correct responses increased as percentage ORM increased.

present study are generally of smaller magnitude.

It was our impression, while running *Ss*, that *Ss* in the 20% ORM group seemed less likely than 100% ORM *Ss* to fail to anticipate an item after once having anticipated it correctly. A score designed to show whether our impression was, in fact, a valid one was computed for all *Ss*. This score, hereafter referred to as the Correct/Opportunity ratio after first correct, was computed as follows. The *S*'s data sheet was inspected to determine when the response member of a given pair of items was first anticipated correctly. Then the number of subsequent opportunities *S* had to respond to the stimulus member of that pair were noted as well as the number of times he responded correctly. This was done for all 10 pairs on the list. Next, the total number correct was divided by the total number of opportunities. The Correct/Opportunity ratio after first correct is the result of this division. Thus, the less frequently *S* misses items after anticipating them correctly once, the higher his score. The results of this analysis are shown in Fig. 1. Inspection of Fig. 1 reveals that

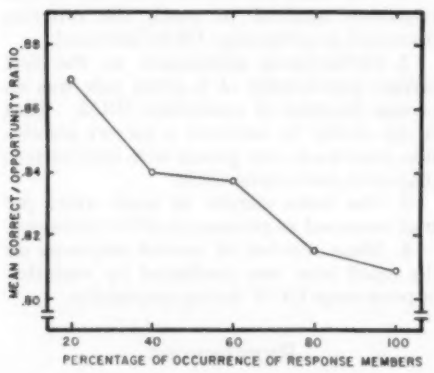



FIG 1. Mean Correct/Opportunity ratios after first correct as a function of percentage of occurrence of response members.

performance in terms of mean Correct/Opportunity ratio after first correct was a monotonic inverse function of percentage ORM. With 20% ORM there were, on the average, 6% more correct responses per opportunity than with 100% ORM. Statistically, the differences among the means of Fig. 1 may be considered reliable ($\chi^2 = 13.8$; $P < .01$). The standard error of these means ranged from .010 to .014.

The mean number of overt errors per trial during acquisition was directly related to percentage ORM. The respective means for 100%, 80%, 60%, 40%, and 20% ORM were 1.09, 1.03, .87, .69, and .58. The differences among these means meet the requirements for statistical significance ($F = 6.22$; $P < .01$). The standard error of these means is .085.

*Retention.*—The mean number of correct responses on the recall trial did not vary as a function of percentage ORM. The means were 6.38, 6.84, 6.58, 6.50, and 6.62 for 100%, 80%, 60%, 40%, and 20% ORM, respectively. The $\sigma_m$ ranged from .24 to .26. Clearly, the null-hypothesis could not be held suspect ($\chi^2 = 1.92$; $P > .80$). Adjustment for possible differences in response strength at the end of acquisition were made via probability analysis (Underwood, 1954), but the conclusion that percentage ORM did not affect retention remained unaltered. The relearning data mirrored the recall data and therefore need not be presented.

## DISCUSSION

*Acquisition.*—The fact that the present differences in performance as a function of percentage ORM were of much smaller magnitude than those obtained by Goss et al. (1959) can most probably be attributed to the difference in the level of meaningfulness of the learning ma-

terial employed in the two studies. This seems especially likely because meaningfulness is known to be a potent source of variance in verbal learning (Underwood & Schulz, 1960). Therefore, it appears reasonable to infer from these studies that the effectiveness of variation in percentage ORM may be inversely related to the meaningfulness of the material to be learned.

The general failure to find more marked decrements in performance with reduced ORM surprised us, as it did Goss et al. The *Ss* in the present 20% ORM group saw the response members of the list an average of 4.5 times in learning to criterion as contrasted with 17.5 exposures of each response member with 100% ORM. Goss has pointed out some factors which may have counteracted the effect of reduction in ORM. The present results seem pertinent to two of these factors; namely, amount of opportunity for rehearsal and amount of response competition per trial.

Consideration of the inverse relationship between percentage ORM and Correct/Opportunity ratio after first correct (cf. Fig. 1) in conjunction with the data on trials to learn to criterion suggests that the detrimental effects of reduced ORM reflected by the latter measure must have been confined to the relatively earlier stages of learning which primarily involved the accomplishment of the first correct anticipation of a given response member, since maintenance of the response thereafter proceeds better with intermittent ORM. The differential opportunity for rehearsal is also greatest during the early stages of learning and decreases steadily as *S* attempts to anticipate more and more response members in the low ORM condition. Therefore, rehearsal is undoubtedly an important factor in the attenuation of the effectiveness of reduced ORM.

Finally, if the occurrence of overt errors can be taken as an index of response competition, then the direct relationship between frequency of overt errors per trial and percentage ORM found in the present study suggests that response competition does in fact increase as percentage ORM increases. Hence, the second important source of compensation for the reduction of ORM must be associated with reduced response competition throughout learning under conditions of intermittent ORM. Unfortunately, the present data fail to provide any clues as to a possible mechanism by means of which intermittent ORM could reduce competition.

*Retention.*—The retention results unequivocally failed to support the hypothesis that intermittent ORM increases resistance to forgetting under the conditions of the present experiment. Perhaps the present experiment was simply a poor one for a test of the hypothesis. A better test might be provided by an experiment in which the traditional A-B, A-C retroaction paradigm was utilized to study amount of retroactive inhibition as a function of percentage ORM in original learning.

## Summary

The effect of variation in percentage of occurrence (20%, 40%, 60%, 80%, 100%) of response members (ORM) on the learning and retention of paired-adjective lists was studied. A total of 250 *Ss*, 50 per list, learned the same list under the five ORM conditions to a criterion of one perfect recitation. Retention was measured 23.5 hr. after acquisition. The following results were obtained:

1. The mean number of trials and correct responses required to reach the criterion decreased as percentage ORM increased.
2. Performance subsequent to the first correct anticipation of a given pair was an inverse function of percentage ORM. That is the ability to maintain a correct association, once made, was greater with intermittent than with continuous ORM.
3. The mean number of overt errors per trial increased as percentage ORM increased.
4. Mean number of correct responses on the recall trial was unaffected by variation in percentage ORM during acquisition.

## References

Dixon, W. J., & Massey, F. J., Jr. *Introduction to statistical analysis.* (2nd ed.) New York: McGraw-Hill, 1957.

GOSS, A. E., MORGAN, C. H., & GOLIN, S. J. Paired-associates learning as a function of percentage occurrence of response members (reinforcement). *J. exp. Psychol.*, 1959, **57**, 96–104.

MELTON, A. W., & IRWIN, J. M. The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. *Amer. J. Psychol.*, 1940, **53**, 173–203.

MELTON, A. W., & VON LACKUM, W. J. Retroactive and proactive inhibition in retention: Evidence for a two-factor theory of retroactive inhibition. *Amer. J. Psychol.*, 1941, **54**, 157–173.

UNDERWOOD, B. J. Retroactive and proactive inhibition after five and forty-eight hours. *J. exp. Psychol.*, 1948, **38**, 29–38.

UNDERWOOD, B. J. Studies of distributed practice: II. Learning and retention of paired-adjective lists with two levels of intra-list similarity. *J. exp. Psychol.*, 1951, **42**, 153–161.

UNDERWOOD, B. J. Speed of learning and amount retained: A consideration of methodology. *Psychol. Bull.*, 1954, **51**, 276–282.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning.* Chicago: Lippincott, 1960.

(Received July 1, 1959)

*Journal of Experimental Psychology*
Vol. 59, No. 6, 1960

# A TEST FOR INTERACTION OF DELAY OF KNOWLEDGE OF RESULTS AND TWO TYPES OF INTERPOLATED ACTIVITY[1]

EDWARD A. BILODEAU AND FRANCIS J. RYAN

*Tulane University*

Bilodeau and Bilodeau (1958) recently published results from five studies on delay of knowledge of results (KR) in motor learning tasks. In none of these five studies was delay a significant variable. Delays by seconds, hour, day, and week produced only null results in lever-pulling and knob-turning tasks. Null results were not surprising, for the responses learned were relatively simple and discrete, no special interfering activities were interpolated, and potentially confounding temporal variables were carefully controlled, i.e., intertrial interval and post-KR interval.

Moreover, it was shown that these results were well supported by the literature. There was, however, one outstanding exception that proved difficult to explain without introducing petty criticisms. This was the study by Greenspoon and Foreman (1956) that not only produced significant results but also yielded a handsome set of delay curves with 0-, 10-, 20-, and 30-sec. parameters. Greenspoon and Foreman's conclusion—"The results of this experiment demonstrate the efficacy of immediate knowledge of results in the learning of a motor task,"—was quite at variance with that of Bilodeau and Bilodeau. It was judged that by far the best way to deal with the conclusion was to hypothesize that the findings could not be replicated.

[1] This work was supported by a grant from the National Science Foundation.

The essence of Greenspoon and Foreman's method was as follows. The *S*s were taught to draw 3-in. straight lines while blindfolded. KR was "long," "short," or "right." There were eight *S*s in a group; each *S* drew 50 lines at 30-sec. intervals. To Greenspoon and Foreman, the critical part of the procedure was that after drawing a line *S* retained his grip of the pencil, kept hand and arm free of the table, and allowed *E* to return the hand to the starting point at the end of the 30-sec. intervals. The foregoing, called "hand-maintaining activity," required *S* to keep his arm off the table for the whole 25-min. testing session and must have been quite tiring.

In an earlier study by Saltzman, Kanfer, and Greenspoon (1955) in which delay of KR had no effect on line drawing, the procedure was different in two respects. First, *S* rested his elbow on the table and briefly raised his pencil when the line was finished. (Again, *E* returned *S*'s hand to the starting point.) Second, the point at which the response originated was only approximately the same from trial to trial. Greenspoon and Foreman allow the possibility that either procedural difference may account for the discrepancy of outcome. As they put it, (*a*) delayed information may differentially reinforce hand-maintaining activity so that longer delays of information are less effective than shorter delays, and (*b*) Saltzman's et al. slightly variable starting point made enough of a stimulus change from trial to trial to reduce the effects of delay.

If we were convinced of the reliability of the above differences in outcome, we might offer arguments about the interpretation, but our basic disagreement is over fact, not interpretation. It is difficult to persuade us, with a single experiment, that such procedural differences as described could make delay of KR effective. The major objective of the

present study, thus, was to replicate two of their groups. Also tested were two groups of *S*s who kept the hand at rest between responses, with the starting point controlled.

## METHOD

*Subjects.*—The *S*s were 48 undergraduate men and women of Tulane University. Men and women were assigned in equal number to each of four groups.

*Apparatus.*—The apparatus was patterned after that of Greenspoon and Foreman: a continuous strip of paper rolled over a piece of plate glass fastened to a laboratory table. The paper slid beneath a carpenter's square. Thus, the glass formed a smooth writing surface beneath the paper and the square served as a straight edge for *S*'s pen. The inner corner of the square was the starting point for the pen and, of course, the rulings on the square permitted *E* to measure the length of line as it was drawn. The apparatus was covered with a black cloth except when *S* was blindfolded.

A Lindy pen was used instead of a mechanical pencil. This substitution improved the riding qualities of the instrument along the square. Furthermore, the properties of the writing point were less variable and the point could not be broken. Adhesive tape was wrapped around the pen from 2 to 2½ in. from the point. The tape made a nonslip grip during the hand-maintaining procedures. The tape also served the useful purpose of standardizing the grip between *S*s and for trials within *S*s.

*Procedure.*—A 2 × 2 factorial combination of delay of KR and maintenance was used. KR was administered almost immediately after a response (I) or was delayed by 20 sec. (D). The hand and arm were maintained (Hold) or they were not (Rest).

The *S*s, blindfolded, had 50 trials. When the first 35 trials were under Treatment I, the last 15 were D trials. When the first 35 trials were under Treatment D, the last 15 were I trials.

The basic sequence of events for all groups was a ready signal, a "draw" signal 5 sec. later, and then KR followed in turn by another ready signal. The draw signals occurred every 30 sec. All four groups were kept to a 30-sec. intertrial interval; the post-KR periods were approximately 30 and 10 sec. long for the I and D treatments, respectively.

KR was "long," "right," and "short." The *E* said "right" when the line was within .25 in. of the 3.00-in. goal. However, *E* recorded *S*'s response error (line length less 3.00 in.) irrespective of KR.

The following practices define hand maintaining and the Rest and Hold treatments.

*Treatment rest.*—Upon completing his line, *S* returned his hand and arm to his lap. The *S* held the pen, but was otherwise relatively relaxed. "Hand" spoken by *E* was the signal to present hand and pen to *E* who then guided them to the starting point. "Hand" preceded "Draw" by 5 sec.

*Treatment hold.*—Upon completing his line, *S* initiated no further responses. He maintained his grip of the pen, kept the pen point where he had stopped it, and kept his arm off the table (forearm parallel to the table top). "Hand" spoken by *E* was the signal that *E* was just about to grasp and to guide *S*'s hand to the starting point. Again, the "Hand" signal preceded the "Draw" signal by 5 sec. Treatment Hold was unpleasant, for *S* had to hold his arm out for about half an hour. Nevertheless, all *S*s followed directions.

*Design compared with Greenspoon and Foreman.*—In all essential respects, Groups I-Hold and D-Hold are replicates of the groups with 0- and 20-sec. delays in the study by Greenspoon and Foreman. One immaterial difference is that our last 15 trials are transfer shift trials. This modification was possible because the differences between their groups were produced during the early trials and not during the late trials.

Had it been possible or desirable, we should, of course, have repeated the condition with the longest delay, i.e., 30 sec. It would seem nearly impossible to have a 30-sec. interresponse interval and delay KR by 30 sec. If the administration were possible, it would be generally undesirable for *S* would have 0 sec. to respond to his feedback.

Greenspoon and Foreman reported the results in terms of right responses. Right responses and the more analytic measure, magnitude of error, will be reported here.

## RESULTS

The data were first examined for the mean number of right responses in successive blocks of five trials, an analysis identical with that of Greenspoon and Foreman.[3] There were no

[3] Tables of data have been deposited with the American Documentation Institute. Order Document No. 6301, remitting $1.25

consistent differences between groups, there being several changes in rank over blocks of trials. An important question is whether the present results are of the same order of magnitude as those of Greenspoon and Foreman. They obtained mean *rights* of 27.38 and 15.00 for Groups I and D, respectively. Projecting our 35-trial results to a base of 50 trials, we obtain 27.74 and 26.67 for Groups I and D, respectively. There is, thus, remarkable agreement for I and marked disagreement for D. Obviously, our D *S*s outperformed theirs. Our D *S*s were about twice as variable, but our I *S*s were comparable. Moreover, within their own experiment, D *S*s showed the lesser variability. It is our guess, then, that the fundamental contradiction between studies is attributable to their D group which is low in mean and variability.

A more analytic measure of the line-drawing response is the mean absolute error. Means based upon successive blocks of five trials are plotted in Fig. 1. The left side of the figure represents the trials before the shift in delay, the right side the trials after the shift. The only discernable trend is that of learning; no differences appear ascribable to delay of KR or to maintenance.

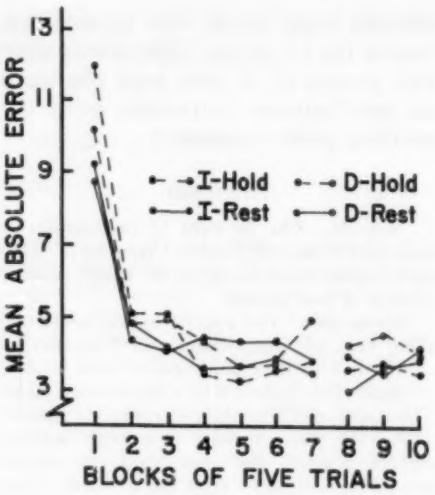



FIG. 1. The effect of delay of knowledge of results and of maintenance. For Blocks 8, 9, and 10, the I groups were switched to Delay and the D groups were switched to Immediate KR.

The means and variances of the totals for *right* and *errors* per *S* over the first 35 trials are given in Table 1. The column means, based upon 24 *S*s, show that the greater number

for 35-mm. microfilm or $1.25 for 6 by 8 in. photocopies.

TABLE 1

MEANS AND VARIANCES FOR THE FIRST 35 TRIALS
(Right represents the number correct; error (1/16 in.) represents the absolute error per *S*)

| Maintenance | Delay | | | | Mean | |
|---|---|---|---|---|---|---|
| | I | | D | | | |
| | Right | Error | Right | Error | Right | Error |
| Hold | 19.42 (12.63) | 185.17 (2088.15) | 18.67 (18.79) | 184.33 (2596.42) | 19.04 | 184.75 |
| Rest | 18.42 (13.72) | 174.83 (648.70) | 19.50 (11.54) | 162.75 (1053.11) | 18.96 | 168.79 |
| Mean | 18.92 | 180.00 | 19.08 | 173.54 | 19.00 | 176.77 |

TABLE 2

MEANS AND VARIANCES FOR THE LAST 15 TRIALS
(Right represents the number correct; error (1/16 in.) represents the absolute error per *S*. Group I has been switched to Delay and Group D has been switched to Immediate KR)

| Maintenance | Former Delay | | | | Mean | |
|---|---|---|---|---|---|---|
| | I | | D | | | |
| | Right | Error | Right | Error | Right | Error |
| Hold | 10.25 (1.66) | 54.75 (148.93) | 8.58 (9.36) | 64.50 (546.82) | 9.42 | 59.62 |
| Rest | 9.17 (6.52) | 58.25 (202.20) | 9.75 (5.30) | 52.92 (267.17) | 9.46 | 55.58 |
| Mean | 9.71 | 56.50 | 9.16 | 58.71 | 9.44 | 57.60 |

of rights and the lesser error were obtained by Treatment D. Neither the Delay nor Maintenance variable, however, exerts much influence for all of the entries are exceedingly similar. This was amply confirmed by analysis of variance of the two factorials in Table 1. Most of the *F*s were less than 1.00 and none approached the .05 level.

The right responses and the errors for Trials 36–50 are shown in Table 2. These are the transfer-shift trials; Treatment I actually received delayed information and Treatment D received immediate information. Once again all of the entries are very similar. In terms of numerical differences, both measures of response favor the treatment with former immediate KR and present delay of KR. The *F* tests, however, failed by a wide margin to reach the .05 level.

There are eight opportunities to compare delay with immediate information in the cells of Tables 1 and 2. In five of the eight, the better performance is associated with the delay treatment. These comparisons are mentioned in order to make a minor point—the difficulty in maintaining that the better treatment is I. The major point is that there is no source of significant variance in Tables 1 and 2.

## DISCUSSION

There was no evidence of special efficacy of immediate KR in the present study. Greenspoon and Foreman give evidence of such efficacy when the hand was maintained. They suggested that the positive result might be obtained when the hand was maintained, but not when the hand was not maintained. We have found no evidence that maintenance, nonmaintenance interacts with delay of KR, and no evidence that delay of KR impairs the learning process. Further, the maintenance variable itself did not produce significant differences.

The effects of delay of KR upon the learning process have been recently discussed by Bilodeau and Bilodeau (1958). Maintenance is another matter. It may or may not interact with delay of KR, depending upon what we mean by the word.

*Meanings of maintenance.*—If *S* suspended activities after drawing a line, the traces of the response would be undisturbed. A complete suspension has never been achieved. It is typical procedure for *S* to initiate and execute a number of responses other than line

drawing. Some of these involve the hand and arm and their movement to the table top, to the starting point, and so on. These extra responses are included within nonmaintenance procedures. As used here, maintenance also refers to certain observable responses of the hand and arm, e.g., the pencil- and arm-holding actions. Maintenance, as Greenspoon and Foreman recognized, was not the opposite of nonmaintenance. Neither the Rest nor Hold techniques used here produced or even approximated a state of suspended activity.

Since the Thorndikian line-drawing task is not especially suited to achieving anything but a moderate degree of control over *S*'s extra responses, we prefer the Rest treatment as standard, neutral procedure. At least at present, *S* must do something with his arm and hand before and after KR, and the hand must be positioned for the beginning of the drawing response. Bilodeau and Bilodeau have suggested that dropping the arm to one's side after responding and then engaging in normal waking activities offer just about as standardizing (and neutral) a situation as we are likely to find for some time. The value of normal activity as a standard procedure is underscored when pre- and post-KR delays are of the order of hours, days, and even years. Maintenance at any level (limb or trace) would seem out of the question with such intervals.

We allowed that both Rest and Hold activities represent potential interfering tasks, though not differential ones. Our null results do not permit concluding that level of interference varies from Rest to Hold, from I to D, and in the interaction of these two variables. But the extent to which the outcome of the present study is restricted by our range of values is most pertinent. As far as delay of KR is concerned, the range of variation seems adequate, Bilodeau and Bilodeau (1958) having shown that delays of KR up to week-long intervals had no adverse effects upon learning. As to the separation between Rest and Hold we are not quite as certain. First, by *S*s introspections, Hold is a lot more tiring than Rest (see the section below). Second, our critique of the procedures Rest and Hold shows that neither is a condition of suspended activity. Third, the Rest-Hold differences in the results were null. Thus, there are two good arguments that Rest is not especially different from Hold.

Someone will no doubt suggest that the second argument above shows the need for a control group which is completely free of interpolated activity. This type of control is not yet technically feasible as is made clear in the debates over the techniques of freezing, hypnosis, and drugs in testing retroactive inhibition theory.

*Effects of Rest and Hold treatments.*—There was no over-all difference between the two conditions of maintenance. This is a puzzle if nonmaintenance (Rest) implies a condition of relative rest between line-drawing responses, and maintenance (Hold) implies a condition of work. Having equated maintenance with a species of massed practice, and nonmaintenance with spaced practice, we might expect performance to favor nonmaintenance. It might be that a condition of massed activity plus less interference has been placed in balance against spaced activity plus greater interference. Though this is a possibility, our preferred interpretation of the null result is that (*a*) the interferences are essentially equal, and (*b*) the differential fatigue exists but is irrelevant *to the accuracy* of line drawing. By analogy, the overtired lathe worker can produce sufficiently accurate pieces of work. Such pieces can be produced despite protestations of fatigue. In the laboratory, the ancient ergographologists often succeeded in producing a good deal of fatigue though little finger-decrement. Similarly, while Hold *S*s may be more fatigued than the Rest *S*s, this does not imply that they cannot draw lines as well. It does imply that there is *some* activity which cannot be performed as well. This "some" activity need not necessarily be the same activity which produced the intervening state. Under the circumstances where

the state does depress the activity that produced the state, the decrement may be in one measure of response and not in another.

## Summary

This experiment replicates two groups of a line-drawing study by Greenspoon and Foreman, i.e., 0- and 20-sec. delay of knowledge of results. Their data gave evidence of the efficacy of immediate knowledge of results. The present experiment gave null results, the outcome being entirely consistent with a number of earlier studies.

The *S*s of the two groups mentioned above followed a "hand maintaining" procedure in that the drawing hand and arm were kept off the table between responses, and in that *E* positioned *S*'s hand for the start of each response. Two other groups, also 0- and 20-sec. delay, did not maintain the responding limb, but returned it to the lap and rested between responses. Neither procedure led to differences between immediate and delayed information.

The maintenance procedure used here was very tiring. Nevertheless, there was no difference between any maintenance and nonmaintenance groups. This was interpreted to mean that fatigue is selective and need not impair the accuracy of all types of responses even over an extended number of trials.

Possible maintenance procedures and how they might influence the trace of the line-drawing response are discussed.

## References

Bilodeau, E. A., & Bilodeau, I. McD., Variation of temporal intervals among critical events in five studies of knowledge of results. *J. exp. Psychol.*, 1958, **55**, 603–612.

Greenspoon, J., & Foreman, S. Effect of delay of knowledge of results on learning a motor task. *J. exp. Psychol.*, 1956, **51**, 226–228.

Saltzman, I. J., Kanfer, F. H., & Greenspoon, J. Delay of reward and human motor learning. *Psychol. Rep.*, 1955, **5**, 139–142.

(Received July 1, 1959)

*Journal of Experimental Psychology*
Vol. 59, No. 6, 1960

# EFFECTS OF SITUATIONAL ANXIETY AT DIFFERENT STAGES OF PRACTICE [1]

JOAN L. BARDACH [2]

*New York University*

The hypothesis investigated in the present experiment is that situational anxiety induced early in practice will have a more unfavorable effect upon performance than anxiety induced late in practice. Insofar as anxiety is viewed as a drive that energizes all activities that may be in progress, the stage in practice at which it is introduced should affect performance. Early in practice, the strength of the correct responses relative to that of the incorrect responses is less than it is later in practice. Indeed, early in practice, especially with difficult tasks, the incorrect responses may be the stronger, and introduction of anxiety therefore may be detrimental to learning and performance. Later in practice, however, the strength of the correct responses may exceed that of the incorrect ones, so that the introduction of anxiety should facilitate performance, or at least be less detrimental than earlier introduction would be.

If we consider anxiety as a class of responses, we can make the same prediction. The responses associated with anxiety possess a certain degree of strength. If these responses are of such a nature that they interfere with the correct responses, the observed effects of this interference will depend upon the relative strength of these two kinds of reactions. In an early stage of practice, the strength of the correct responses is less than it is in a later stage. Hence the amount of impairment of performance produced by interfering responses associated with anxiety should be less in later stages of practice. In making this prediction, it is assumed that the strength of the anxiety reactions does not increase markedly during learning.

## METHOD

*Subjects.*—The *Ss* for this experiment were volunteer, undergraduate men from New York University. Only male *Ss* were used in order to eliminate sex differences in susceptibility to electric shock.

*Learning material.*—The learning material was a list of 10 nonsense syllables, selected from Glaze's list (1928) and of 26.67% association value. Nonsense syllables of low association value were used because previous experimental results indicated that such syllables present a learning task sufficiently difficult for learning to be impaired either by high anxiety, as measured psychometrically (Taylor, 1953), or by situational anxiety induced by unavoidable shock.

*Apparatus.*—Each nonsense syllable was exposed for 2 sec. on a Patterson memory drum, Model 1B. The intertrial interval was 20 sec. Shock was administered by means of a Porter inductorium adjusted for tetanizing shock; i.e., the vibrator make-break mechanism was used. The primary coil and make-break vibrator were in series with a telegraph key and one No. 6 dry cell.

The electrodes were pieces of aluminum

[1] Based on a dissertation submitted in partial fulfillment of the requirement for the PhD degree at New York University. The author gratefully thanks Leland Crafts, Research Advisor, whose constructive ideas and supportive attitudes contributed so materially to this research. The author also thanks Ralph Gilbert and Robert Silverman, members of her Thesis Committee, for their helpful comments in which problems in method and theory were anticipated.

[2] Now at the Institute of Physical Medicine and Rehabilitation.

mounted side by side on a wooden block, against which *S*'s fingers were strapped. Leads to the electrodes were taken from the secondary terminals of the induction coil.

*Procedure.*—The *S* was asked to anticipate each syllable by spelling it. In accordance with the procedure of Silverman and Blitz (1956), all *S*s were given a series of preliminary training trials in which they were required to learn a list of six simple three-letter words that were presented in the same manner as the experimental task. Following the preliminary training, each *S* learned the experimental list of nonsense syllables to the criterion of two consecutive errorless trials.

Four groups of *S*s, 17 per group, were used. At the prescribed stage of practice, *S* in an experimental group was suddenly told: "From now on you will receive electric shocks at any time." While these instructions were being given, electrodes were attached to *S*'s fingers. The total time required for the verbal instructions and for attaching the electrodes was 30 sec. At the conclusion of the trial following this procedure, one shock actually was given. However, no shock was given at any later time, although presumably *S* continued to expect one.

Shock was administered in the experimental groups either after *S* had learned 30% of the list, i.e., had anticipated three syllables correctly (Group E3) or after *S* had learned 70% of the list, i.e., had anticipated seven syllables correctly (Group E7). Group C3, the control for Group E3, was equated with the latter in terms of the number of trials to reach the criterion of 30% correct anticipations on the experimental list. Group C7, the control for Group E7, was equated with the latter in terms of the number of trials to reach criterion of 70% correct anticipations on the experimental list. Since the two experimental groups had an interval of 30 sec. following the learning of 30% or 70% of the list, respectively, their control groups had an interval of the same length following the achievement of the same criteria. In order to prevent rehearsal of the rote-memory task by the control *S*s, they were asked simple questions, such as age, college major, etc., during this 30-sec. interval.

After completion of the learning task, each experimental *S* filled out a questionnaire designed to reveal his reactions to shock-warning and to shock-administration. At the conclusion of the experiment, all *S*s were given an explanation of the experiment and attempts were made to allay whatever anxiety might have been aroused.

## Results

Results of the analysis of variance, using mean total trials to learn to criterion, are shown in Table 1. These data indicate no statistically reliable differences in performance among the groups as a whole, nor between the groups that were given shock and those that were not, nor between those groups that were interrupted early in learning and those that were interrupted late. Further study of the interaction between time of interruption and shock—no shock again revealed no differences. Comparisons of trials or errors to criterion for the various groups by *t* tests (matched groups formula used where appropriate; two-tailed test of significance) revealed only one significant ($P < .05$) difference between the means. Between Groups E7 and C7, the *t* was 2.68 ($P = .02$, $df = 16$). This finding indicates that it is the late introduction of anxiety that results in impaired performance, rather than its early introduction as was the theoretical expectation. This result does not conform to predictions based on a view of anxiety either as a drive or as a class of interfering responses.

However, qualitative data noted during the running of the experiment, as well as the results of the question-

TABLE 1

Analysis of Variance of Trials to Criterion

| Source | *df* | *MS* | *F* |
|---|---|---|---|
| Between pairs | 33 | | |
| B (7 vs. 3) | 1 | 187.78 | 1.59 |
| Error (b) | 32 | 117.77 | |
| Within pairs | 34 | | |
| A (E vs. C) | 1 | 101.31 | .90 |
| AB | 1 | 106.25 | .94 |
| Error (w) | 32 | 112.44 | |
| Total | 67 | 115.85 | |

TABLE 2

NUMBER OF TRIALS TO LEARN TO CRITERION

| Groups Compared | *N* | *Mean* | *SD* | *t* |
|---|---|---|---|---|
| E3 Compat. | 6 | 24 | 10 | 2.61* |
| C3 Compat. | 6 | 39 | 13 | |
| E3 Incompat. | 11 | 34 | 9 | 2.32* |
| C3 Incompat. | 11 | 28 | 13 | |
| E7 Compat. | 6 | 29 | 12 | .76 |
| C7 Compat. | 6 | 25 | 10 | |
| E7 Incompat. | 11 | 31 | 9 | 2.71* |
| C7 Incompat. | 11 | 23 | 9 | |
| E3 Compat. | 6 | 24 | 10 | 2.51* |
| E3 Incompat. | 11 | 34 | 9 | |
| E7 Compat. | 6 | 29 | 12 | 1.38 |
| E7 Incompat. | 11 | 31 | 9 | |
| E3 Compat. | 6 | 24 | 10 | .74 |
| E7 Compat. | 6 | 29 | 12 | |
| E3 Incompat. | 11 | 34 | 9 | 1.04 |
| E7 Incompat. | 11 | 31 | 9 | |

* $P = .05$.

naire and a brief terminal interview given the experimental *S*s, suggested that the attitudes induced by the introduction of shock were not the same for every *S*. Failure to take this fact into account in the statistical treatment of the data may explain the negative results obtained.

Therefore, each experimental group was divided into two subgroups, based upon revealed differences in the attitudes induced by shock introduction. One subgroup was composed of *S*s who claimed they were not disturbed by the shock or who believed, despite the instructions, that they would be shocked for *errors* and therefore may have been spurred to greater effort to learn, or of those who felt that by diligent effort they could complete the experiment quickly and thus avoid further shock, or who did not continue to expect shock throughout most of the learning. This group is called the *Compatible* group because the attitudes induced by the experimental conditions seemed to be compatible with the requirements of the learning task. The designation E3 Compat. and E7 Compat. refer to the experimental *S*s who were given shock early or late in the learning, respectively, and in whom compatible attitudes were induced. Groups C3 Compat. and C7 Compat. are comprised of those control *S*s who matched the experimental *S*s in performance before the introduction of the experimental conditions.

The second subgroup into which the experimental groups were divided consisted of those *S*s who were disturbed by the shock, or who felt that shock would be administered randomly, or who continued to expect shock throughout most of the learning. This group is called *Incompatible* because the attitudes induced by the experimental conditions seemed to be incompatible with the requirements of the learning task. The designation E3 Incompat. refers to the experimental *S*s who were given shock early in the learning and in whom incompatible attitudes were induced; C3 Incompat. refers to their control group. E7 Incompat. and C7 Incompat. are the corresponding designations for the experimental group that was given shock late in the learning and its control.

Table 2 shows comparisons of performance when the compatibility or incompatibility of responses associated with the secondary anxiety drive are taken into account. As the table shows, *S*s in the compatible group who received shock early in practice, learned significantly faster than their controls, but *S*s in the compatible group who received shock late in practice did not.

Performance data for the incom-

patible groups are not simply the reverse of those obtained for the compatible groups. Table 2 shows that the experimental *Ss* in whom incompatible responses were induced learned significantly slower than their controls, regardless of whether shock was introduced early or late in practice. The evidence is that Group E3 Incompat. did not take significantly more trials to learn to criterion than did Group E7 Incompat. However, because of the small *N* of 11 in each group, further research is necessary to establish this point definitely.

In order to analyze the course of the effect, or of the immediate and the subsequent effects, of shock introduction, apart from its total over-all effect upon *N* of trials (and of errors) to complete the learning, various learning curves[3] were drawn. Inspection of the curves showing the mean number of correct responses for each trial for the subgroups who received shock *early* in the learning discloses that, after the introduction of the experimental conditions, the performance of the experimental groups in relation to their controls is reversed. After shock, Group E3 Incompat. acquires additional syllables *more slowly* than its control, at least for the subsequent 12 trials, but, after shock, Group E3 Compat. acquires additional syllables *faster* than its control.

Curves were also drawn to show the mean number of trials necessary to learn 1, 2, 3 . . . 10 syllables, and to criterion. Inspection of these curves for the subgroups who received shock *early* in the learning discloses that, after introduction of the experimental conditions, the performance of the experimental groups in relation to their controls is again reversed. Group E3 Compat., throughout the remainder of the learning, acquired each syllable subsequent to shock in *fewer* trials than its control, but Group E3 Incompat., throughout the remainder of the learning, acquired each syllable subsequent to shock in *more* trials than its control.

The comparable curves for the subgroups who received shock *late* in the learning show that the experimental conditions had little effect on performance of those individuals in whom compatible attitudes were induced. It may be that there was insufficient time for these attitudes to take effect. However, these data show that when incompatible attitudes are induced even late in the learning, the acquisition of subsequent syllables is impaired.

[3] These curves can be found in the original dissertation.

## Discussion

The data for the groups as a whole before their division into subgroups in accordance with the compatibility or incompatibility of the induced responses indicate that only late introduction of anxiety impaired performance. If these findings had been the only ones available, the results of the experiment would have been negative, and moreover, would have been difficult to explain in terms of any current theoretical position.

However, when the experimental groups are classified according to the type of induced response associated with the introduction of shock, differences in performance appear that seem more understandable. Now, if the responses to anxiety are compatible with the learning task, early introduction of anxiety is facilitating, although late introduction remains ineffective. If the responses are incompatible, both early and late introduction of anxiety is impairing, although no difference in the degree of impairment is demonstrated. These results support the rather obvious notion that if anxiety is associated with incompatible (interfering) responses, it

will impair learning and performance, but they also indicate that if anxiety is associated with compatible responses, it may significantly facilitate learning and performance, even when task difficulty is held constant.

However, one problem in the interpretation of the results remains, namely, that when anxiety is introduced late rather than early in practice, even those *Ss* in whom compatible responses were evoked were inferior to their controls. A possible explanation of this finding is as follows. In a situation in which compatible responses are induced, a time-lag probably occurs because time is required for the compatible responses associated with the anxiety to produce their potential effects on performance. Table 2 shows that the difference in performance between Groups E7 Compat. and its control is not significant. It may be that the introduction of shock in this experimental group occurred too late in learning for the effects of compatible attitudes to be reflected in performance.

The present results suggest that anxiety can best be conceptualized as having both drive properties and associated responses, a view that is similar to that of Mandler and Sarason (1952). If accurate predictions of performance are to be made, both of these aspects must be taken into account. The data suggest, in addition, that responses associated with the introduction of shock are different for different *Ss*, so that accurate predictions require that these individual differences also be considered. However, since the conclusions from the present study are based on *post hoc* considerations, they must be regarded as tentative until cross-validating studies are made.

## Summary

An experiment tested the hypothesis that situational anxiety induced early in practice will have a more unfavorable effect upon performance than anxiety induced late in practice. Four groups of *Ss* learned a difficult list of nonsense syllables. One group was given unavoidable electric shock after they had learned 30% of the list; another group was given unavoidable shock after they had learned 70% of the list. Each experimental group was matched on preshock performance with a control group.

The data showed that it was late rather than early introduction of situational anxiety that impaired performance, indicating a need for modification of the original hypothesis. When each experimental group was divided into two subgroups in accordance with *Ss'* reported attitudes toward shock, it was found that the different responses induced by the shock had significant effects on performance. The results suggest that when shock induces responses that are compatible with the learning task, facilitation of performance may occur, but that when shock induces responses that are incompatible with the learning task, performance is impaired regardless of when in the course of learning shock is introduced.

## References

Glaze, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–269.

Mandler, G., & Sarason, S. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 166–174.

Silverman, R. E., & Blitz, B. Learning and two kinds of anxiety. *J. abnorm. soc. Psychol.*, 1956, **52**, 301–303.

Taylor, J. A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

(Received July 2, 1959)

*Journal of Experimental Psychology*
Vol. 59, No. 6, 1960

# ROLE OF POSTURAL EXPERIENCES IN PROPRIOCEPTIVE PERCEPTION OF VERTICALITY[1]

RICHARD G. PEARSON

*Carnegie Institute of Technology*

AND GEORGE T. HAUTY[2]

*School of Aviation Medicine, USAF*

In a previous study (Pearson & Hauty, 1959) concerned with the function of learning in the proprioceptive perception of verticality, it was found that alternation of the direction of tilt on successive trials effected a progressive reduction in the error of perceived verticality. In accounting for this effect, two possibilities were discussed: the learning of additional and unique proprioceptive cues to be gained from alternation and the nullification of effects of sensory adaptation which also could have been provided by this experience.

The present study was designed for the purpose of determining the effects of adaptation imbalance and of increase in the delay at tilt upon the error of perceived verticality. Should these effects attenuate reduction in error, sensory adaptation may be regarded as the predominant factor. If not, then support for proprioceptive learning will have been realized.

## METHOD

*Apparatus.*—The tilting-chair used in this study has been described in detail elsewhere (Pearson, 1957). The seat was adjustable to the sitting height of *S* so that the medial axis at eye level would be approximately at the point of chair rotation. The *S*'s control panel contained adjustable, leather sleeve cuffs to support his forearms at positions of tilt, and a three-position switch providing him with directional control of the chair which, of course, was also under the independent control of *E*. The test room itself was lightproof and airconditioned. Both red and white light were available to *E* at his discretion.

*Design.*—The experimental variables were of the following nature: (*a*) Alternation, or directional sequence of tilt, within each of the two magnitudes of tilt: $A_1$—no alternation, in which *all* trials involved the tilting of *S* to his left; $A_2$—alternation, in which *S* was tilted to the left, either 10° or 30°, on all odd-numbered trials and 10° to the right on *all* even numbered trials. (*b*) Tilt magnitude: $T_1$—10° left; $T_2$—30° left. (*c*) Delay at tilt: $D_1$—no delay; $D_2$—30 seconds delay; $D_3$—60 seconds delay.

The *S*s were assigned in equal numbers and randomly to each cell of the $2 \times 2 \times 3$ factorial design. The experimental sample was obtained from a population of volunteer, experimentally naive, basic Air Force trainees. Four *S*s were tested per day until a sample of 96 was attained.

*Procedure.*—The *S* was secured in the tilting-chair, then was indoctrinated in his task. He was told that the test session would consist of 20 trials. If he had no questions, testing then commenced. The *E* moved *S* to a position of tilt, then, following the delay period (if any), transferred control of the chair to *S* who made his judgment. When satisfied with his position, *S* responded "OK" or "All right." The *E* then recorded the direction and amount of error. Thereafter, *E* proceeded in accordance with the "alternation" condition to which *S* was assigned. Delay at the perceived vertical was 20 sec. in all cases.

[1] The contents of this manuscript reflect the personal views of the authors and are not to be construed as a statement of official Air Force policy.

[2] Now at the Civil Aeromedical Research Institute, FAA, Oklahoma City.

TABLE 1

ANALYSIS OF VARIANCE OF ABSOLUTE ERRORS

| Source | *df* | *MS* | *F* |
|---|---|---|---|
| Alternation | 1 | 335.48 | 4.75* |
| Tilt | 1 | 3613.45 | 51.16** |
| Delay | 2 | 20.74 | .29 |
| A × T | 1 | 39.81 | .56 |
| A × D | 2 | 76.14 | 1.08 |
| D × T | 2 | 135.37 | 1.92 |
| A × T × D | 2 | 103.33 | 1.46 |
| Error (a) | 84 | 70.63 | |
| Measures | 9 | 9.83 | 2.42* |
| M × A | 9 | 6.54 | 1.61 |
| M × T | 9 | 8.66 | 2.13* |
| M × D | 18 | 2.62 | .64 |
| M × A × T | 9 | 5.17 | 1.27 |
| M × A × D | 18 | 3.53 | .87 |
| M × D × T | 18 | 2.30 | .56 |
| M × A × T × D | 18 | 4.87 | 1.20 |
| Error (b) | 756 | 4.07 | |
| Total | 959 | | |

* $P < .05$.
** $P < .01$.

## RESULTS

In accordance with the main interests of this paper the statistical analysis was confined to the absolute magnitude of error in perception of the vertical for the odd-numbered trials only. Results of a repeated-measurements analysis of these data are shown in Table 1. Of particular concern to the authors is the finding of significant effects attributable to both the Alternation and Tilt variables. As seen in Fig. 1 the alternation subgroup (10° R), as compared with the "control" subgroup, demonstrates a marked reduction in error beginning at Trial 3, and maintains this superior performance throughout testing. Figure 2 plots the effects of magnitude of tilt upon perception of the vertical, and, as one would expect, the 30°-tilt subgroup has a much larger error of adjustment than the 10°-tilt subgroup. It should be noted at this point that, although both Alternation and Tilt main effects are significant, neither the simple interaction, A × T, nor the second-order interaction, M × A × T, approach significance. The importance of these findings are discussed below.

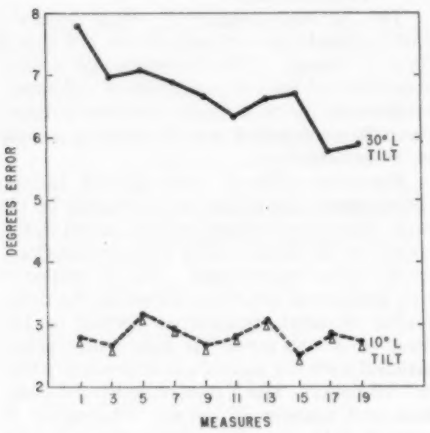



FIG. 1. The effects of alternation upon perception of the postural vertical.

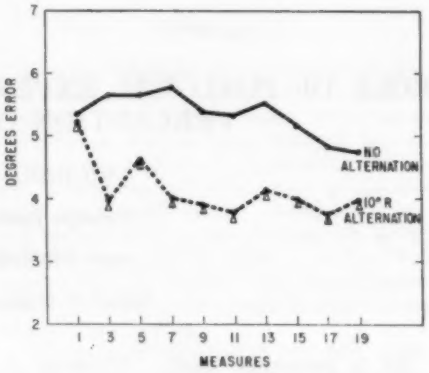



FIG. 2. The effects of magnitude of tilt upon perception of the postural vertical.

The factor of Delay has no significant effect. As shown in Table 1, the Delay *F* ratio is not significant, and, furthermore, none of the interactions involving the Delay variable

with the other main variables are significant. The absolute mean errors of adjustment for the 0-, 30-, and 60-sec. delays were 4.64°, 5.05°, and 4.59° respectively.

The two significant *F*s, Measures and M × T, are doubtless attributable to the progressive reduction in error achieved by the 30°-tilt subgroup (see Fig. 2).

## Discussion

The primary concern of the present study was with the effect of alternation upon perception of the postural vertical. On odd-numbered trials all *S*s were tilted to the left; on even-numbered trials half of these *S*s were tilted a small amount past the true vertical to the right while the other half were tilted again to the left. It was then demonstrated that perception of the vertical by the alternation subgroup was significantly superior to that of the control subgroup. Further, since the interactions, A × T and M × A × T were found not to be significant, it seems reasonable to assume that adaptation imbalance did not attenuate progressive reduction in error of perceived verticality. Similarly, since neither the F ratios for the Delay variable nor those involving interaction of Delay with the other main variables proved to be significant, one must conclude that increased delay at tilt also did not produce such attenuation. Such findings lead the authors to conclude that proprioceptive learning, rather than sensory adaptation, is the determining factor in progressive reduction in the error of perceived verticality.

The confirmation of our previous finding (Pearson & Hauty, 1959) that "passing through the vertical" is an important variable in perception of postural verticality may have important implications regarding the results of work by Mann, Berthelot-Berry, and Dauterive (1949) and Solley (1956). They employed the practice of moving *S* to the opposite quadrant before moving him to the trial position as a precaution against the possibility that *S* would count himself back to the true vertical. It is now clear that this procedure plays a definite and positive role in reducing errors of perception of verticality. Consequently, it is questionable whether the "true" effects of such variables as delay at tilt and rate of chair movement have been isolated in studies wherein alternation has been employed as a precautionary measure.

In agreement with results reported by Clegg and Dunfield (1954) is the finding that delay did not have a significant effect upon perception of verticality. This confirmation supports a previous conclusion (Pearson & Hauty, 1959); namely, that averaging the results of several consecutive trials does not provide a clear picture of the state of performance in these studies. Performance curves may be expected to reveal that with repeated testing certain variables do not have a significant effect upon performance.

## Summary

Ninety-six *S*s were tested on a lateral-tilt chair under darkroom conditions. The task consisted of 20 trials and required *S* to return himself from a position of tilt to the perceived vertical. It was found that neither the procedure of effecting an adaptation imbalance by tilting *S* 30° L, then 10° R, on alternate trials, nor that of increasing the delay at tilt to 30 or 60 sec. attenuated progressive reduction in error of perceived verticality. This would suggest proprioceptive learning, rather than sensory adaptation, to be the factor of principal importance.

## References

Clegg, W. C., & Dunfield, N. M. Non-visual perception of the postural vertical: II. Lateral plane. *Canad. J. Psychol.*, 1954, **8**, 80–86.

Mann, C. W., Berthelot-Berry, N. H., & Dauterive, H. J., Jr. The perception of the vertical: I. Visual and non-laby-

rinthine cues. *J. exp. Psychol.*, 1949, **39**, 538–547.

PEARSON, R. G. The effects of motion-sickness preventives upon orientation in space. USAF Sch. Aviat. Med., November 1957, Rep. No. 58–7.

PEARSON, R. G., & HAUTY, G. T. Adaptive processes determining proprioceptive perception of verticality. *J. exp. Psychol.*, 1959, **57**, 367–371.

SOLLEY, C. M. Reduction of error with practice in perception of the postural vertical. *J. exp. Psychol.*, 1956, **52**, 329–333.

(Received July 3, 1959)

*Journal of Experimental Psychology*
Vol. 59, No. 6, 1960

# SUPPLEMENTARY REPORT: THE YERKES-DODSON LAW AND SHIFT IN TASK DIFFICULTY[1]

VICTOR H. DENENBERG AND GEORGE G. KARAS[2]

*Purdue University*

Broadhurst (1957) has reported a confirmation of the Yerkes-Dodson Law by use of an underwater Y-maze in which motivation was manipulated by the amount of oxygen deprivation prior to the release of *S*, and task difficulty was varied by changing the relative difference in brightness between the two arms of the Y. He also reported that emotional *Ss* exhibited a higher drive level than nonemotional *Ss*. We used Broadhurst's technique to study the effects of infantile handling since it has been suggested that such handling modifies *S*'s emotionality. We report here a pilot study using a straightway as well as some of the Y-maze data of our control *Ss* from the infantile experience study. These groups may be considered to provide a partial replication of Broadhurst's study with different aged *Ss* and with somewhat different delay intervals.

*Method.*—A replica of Broadhurst's maze was used. It was illuminated by two 100-w. bulbs. A black wooden block at the end of the incorrect alley served as the barrier. The correct alley had no barrier. The difference in alley illumination was determined by a Western illumination meter, No. 756, with filters removed. The difference was in the ratio of 1.50.

In a pilot study 5 trials per day were given 21 Harvard Wistar rats (55–61 days old) in a 4-ft. underwater straightway. Time delays of 0, 3, 6, 9, 12, 15, and 18 sec. were used in a 7 × 7 latin square which balanced test days and test sequences. Swimming time was recorded in .01 sec. In the Y-maze study, time intervals of 0, 3, 6, and 12 sec. were used, with 14 *Ss* randomly assigned to each condition. Starting at 51 days *Ss* were given 5 trials a day for 10 successive days. The time in .01 sec. from the moment that the startbox was opened until *S*'s head broke water was recorded, as was the number of errorless trials per day.

*Results.*—The pilot study yielded significant differences ($P < .01$) among Days and Delay Conditions in an analysis of variance of swimming times. The plot of time against days is similar in form to Broadhurst's Fig. 1. The relationship between swimming time and delay time is highly irregular, but increased time tends to be associated with longer delay intervals.

The Y-maze *Ss* averaged 86.6 errorless trials out of 100. This is similar to Broadhurst's "easy" discrimination task where he found a mean of 86.4 errorless trials out of 100. Analyses of the log time scores and number of errorless trials failed to yield significant differences among the motivational levels. However, a significant Days × Motivation interaction ($P < .01$) was found for the time data. The errorless trial data showed that *Ss* had virtually learned the discrimination by the end of Day 5, making 97.3% correct responses during Days 6–10. Thus behavior during Days 1–5 reflects *acquisition* of the response; behavior during Days 6–10 reflects *performance* of the acquired response. The swimming times were separately analysed

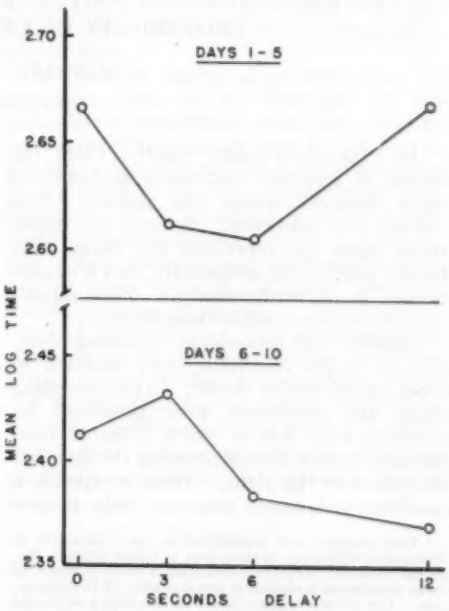



FIG. 1. Swimming time as a function of delay interval for Days 1–5 and 6–10.

[1] This study was in part supported by a grant from the Purdue Research Foundation and in part by Research Grant M-1753 from the National Institute of Mental Health of the National Institutes of Health, United States Public Health Service.

[2] Presently at Department of Psychology, Iowa State University.

for each 5-day period as a function of delay interval. Analyses of these curves (Fig. 1) yielded a significant ($P < .01$) quadratic (curvilinear) effect between swimming time and delay time during Days 1–5 and a significant ($P < .01$) linear function during Days 6–10.

*Discussion.*—Broadhurst, in his straightway study, tested *Ss* first at the 0-sec. delay, then at the 5-sec. delay, etc., thus confounding practice effects and delay intervals.[3] Our data, which separately evaluated these two variables, are not consistent with his, but neither are they consistent with our Y-maze results. It appears that the use of a design in which *Ss* are tested under different delay times is not valid because of differential transfer effects.

Our light–dark ratio was approximately 1:50 while Broadhurst's was about 1:300 for his "easy" task, but the two studies report almost the same percentages of correct responses. However, Broadhurst's "moderate" task had a light–dark ratio of 1:60 (83.1% correct responses); and our data should have come closer to approximating the findings of his "moderate" rather than his "easy" task. The failure to do so can probably be accounted for by the ages of *Ss* involved (213 or 322 days vs. 51 days) and strain differences.

Even though the main effect of Motivation was not significant, our Y-maze results, like Broadhurst's, confirm the Yerkes–Dodson Law when task difficulty is changed by repeated experience. The task changed markedly in difficulty between Trials 1–5 and 6–10. According to the Yerkes-Dodson Law, optimal motivation should be less during the more difficult learning trials (Trials 1–5) than during the latter trials. This effect was obtained, and suggests that *change in task difficulty* as a function of repeated practice must be taken into account when testing the Yerkes-Dodson Law.

[3] P. L. Broadhurst. Personal communication, September 1958.

## REFERENCE

BROADHURST, P. L. Emotionality and the Yerkes-Dodson Law. *J. exp. Psychol.*, 1957, **54**, 345–352.

(Received September 23, 1959)

*Journal of Experimental Psychology*
Vol. 59, No. 6, 1960

# SUPPLEMENTARY REPORT: FURTHER FACTORS AFFECTING THE PROBABILITY OF CHANGING RESPONSES[1]

MAYNARD W. SHELLY II

*Office of Naval Research*

In a previous study (Shelly, 1958) the effects of response contingent probabilities upon response change was studied. Two factors were identified: changes are conditional upon (*a*) repetition not being reinforced, and (*b*) the probability that a change results in a reinforcement. This supplementary report extends these results.

*Method.*—All procedures, including numbers of *Ss* per condition, were identical to those used in Shelly (1958). In the preceding study the conditions were presented by matrices, each row of which differred from the one above it through moving the elements one place to the right. Thus, to specify a condition it becomes necessary only to give the first row with the understanding that the first position specifies the probability that repetition will be reinforced. Using this method of presentation, five nonexclusive sets of conditions can be described: Set A: {(.00, 1.0, .00, .00), (.00, 1.0, 1.0, .00), (.00, 1.0, 1.0, 1.0)}; Set B: {(.00, .33, .33, .33), (.00, .50, .50, .00), (.00, .50, .50)}; Set C: {(.50, .00, .00, .00), (.50, .25, .00, .00), (.50, .50, .00, .00), (.25, .25, .25, .25), (.50, .50, .50, .50)}; Set D: {(.50, .50, .00, .00), (.50, .25, .25, .00), (.50, .16, .16, .16)}; and Set E: {S = (.75, .25, .00, .00), T = (.25, .75, .00, .00), Z = (.00, .75, .25, .00)}. For Set E, the labeling of conditions is done upon the basis of the first letter of the percentage of reinforcement associated with repetition, e.g., S for 75%. To refer to any other condition, e.g., the first condition of Set A, the notation $A_1$ will be used.

*Results and conclusions.*—Within the sets of conditions, methods of analysis identical to those used in the previous paper, including

[1] This research was completed at the University of Wisconsin while the author was a Social Science Research Council Research Training Fellow. Some of these results are included in the author's PhD dissertation. The author wishes to thank the Office of Naval Research for providing him with time, and the Department of Psychology, Northwestern University for the facilities to complete this research.

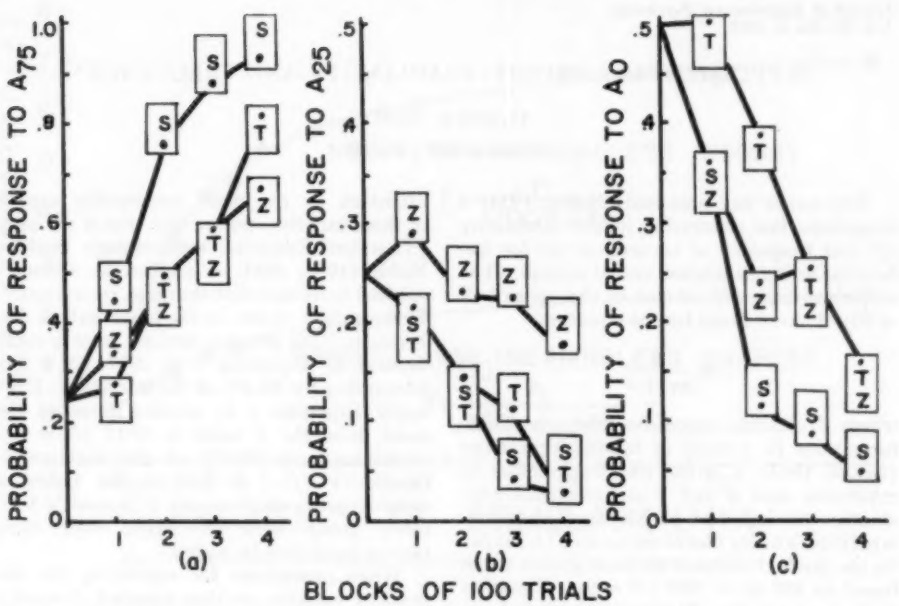



FIG. 1. Probability of responding to an alternative when repetition is reinforced with varying probabilities.

a constant confidence level of .05, will be used. "$P_C$" stands for the probability of changing responses on successive trials. To indicate that, for example, $P_C$ for $A_1$ and $A_2$ was greater than $P_C$ for $A_3$ on Blocks 3 and 4 of 100 trials, the short hand notation $(P_C(A_1, A_2) > P_C(A_3) | 3, 4)$ will be used.

The following are the conclusions together with the supporting results.

1. Increasing the reinforcement of repetition, even though changing responses leads to equal reinforcement, reduced $P_C$: $(P_C(C_4) > P_C(C_5) | 1, 2, 3, 4)$.

2. A larger reinforcement for a specific change appears to be superior in increasing $P_C$ than the same probability spread over a number of alternatives: $(P_C(D_1) > P_C(D_2) | 2)$.

3. The differences in (2) were small, however, and when the number of alternatives to which a change can be made is increased *without* any decrease in the probability that a change can lead to a reinforcement, then $P_C$ may increase: $(P_C(A_3) > P_C(A_1) | 1)$.

4. In agreement with Conclusion 2, when the probability of a single change leading to a win is increased *without* altering the probability of reinforcing repetition, $P_C$, if it changes significantly, increases: $(P_C(C_3) > P_C(C_2, C_1) | 2)$; $(P_C(C_3) > P_C(C_1) | 4)$.

5. There was some evidence that number of available alternatives had an initial effect on $P_C$. Although the only significant relation was $(P_C(B_1) > P_C(B_2) | 1)$, $P_C(B_2)$ lay regularly between the other two corresponding probabilities.

6. That the choice of an alternative on successive trials can be influenced by the value of the conditional probability of reinforcing repetition can be seen in Fig. 1. In this figure $A_{75}$ refers to that alternative which on a given trial has the probability of reinforcement of .75 associated with it; corresponding interpretations apply to $A_{25}$ and $A_0$. Where two values do not differ significantly for the same block of 100 trials, they are combined into a single point. If the probability of reinforcement were independent of changing or repeating responses, all three graphs would exhibit but a single curve.

The preceding results extend the findings of Shelly (1958) and would seem to indicate that factors influencing changing responses need to be considered in predicting rate of learning.

## REFERENCE

SHELLY, M. W. The effects of response contingent probabilities which favor response change. *J. exp. Psychol.*, 1958, **56**, 239–245.

(Received February 10, 1960)

*Journal of Experimental Psychology*
Vol. 59, No. 6, 1960

# SUPPLEMENTARY REPORT: FAMILIARITY AND FREQUENCY[1]

CLYDE E. NOBLE
*Montana State University*

The writer has proposed (Noble, 1954) a hyperbolic law connecting judged familiarity ($f$) and frequency of occurrence ($n$) for unfamiliar and meaningless verbal stimuli. The empirical data, reproduced in the upper half of Fig. 1, were fitted by the equation,

$$f = \frac{n}{an + b} \qquad [1]$$

where $f$ denotes successive-intervals transformations in $z$ units of familiarity ratings (Noble, 1953), $n$ is the physical number of exposures, and $a$ and $b$ are empirical constants. An index of correlation ($p$), calculated from a linear plot of reciprocals ($1/f$ a $1/n$) by the product-moment method, gave a value based on 288 *S*s of .998 ($N = 9$). Equation 1 thus accounts for 99.6% ($p^2 \times 100$) of the variance in the dependent variable.

Repeating this experiment with random visual shapes on 600 *S*s from the same population, but using a graphical scaling method, Arnoult (1956) obtained a similar curve. To describe it, however, he chose the logarithmic function,

$$f = a \log (n + b) - c \qquad [2]$$

Arnoult reported a higher index of correlation for a semilog reduction of Equation 2 ($p = .997$), in which $b = 1$, than for the linear plot resulting from taking reciprocals of Equation 1 ($p = .950$).[2] Comparing $p$ values from reductions of Equations 1 and 2, however, is dubious practice because mathematically different transformations are involved. One should not forget that $p^2$ then describes the precision of the general law modified by scale characteristics.

Goodness of fit depends, furthermore, upon the particular scaling operations followed in calculating familiarity scores and upon whether any *ad hoc* constants are added to the independent variable $n$. Limiting our attention to the most comparable aspects of these experiments, we find that if the same (Thurstone-Edwards) methodology used by Noble (1953, 1954) is applied to Arnoult's original frequency distributions, the rectangular hyperbola shown in the lower half of Fig. 1 results. Its $p$ value derived from a rectification of Equation 1 is .997 ($N = 10$), accounting for 99.4% of the variance. If we apply Equation 2 in semilog form to the same data, the $p$ value is .9917 ($N = 10$), explaining only 98.3% of the variance in familiarity ($f$). A still smaller index of determination might result if Arnoult's 5-hr. delay group were eliminated since their ratings tended to be higher.

When procedures for measuring the dependent variable are thus equated, Arnoult's data confirm Noble's finding that familiarity ($f$) is determined almost uniquely by frequency of exposure ($n$), and that this dependency grows according to a hyperbolic law. The precision and generality of Equation 1 is striking; only the slope ($b$) and intercept ($a$) parameters of the reduction line, $1/f = b(1/n) + a$, change when the task and stimulus characteristics are experimentally altered. Errors of measurement are reduced to the order of 1% in both cases.

Because of the "extraordinary accuracy" of the empirical fits of Equations 1 and 2, a casual reader may be tempted to conclude that these two sets of observations are "genuinely different" (Arnoult, 1956, p. 275). After descriptively asserting that "$f$ is a monotonic, negatively accelerated function of $n$" (p. 272), Arnoult proceeded to embrace the notion of an infinitely-rising curve (p. 275) which is conventional among perception researchers. Hyperbolas have asymptotes, but logarithmic curves are unbounded. On the hyperbolic hypothesis, the calculated asymptote ($1/b$) for Noble's data has a maximum $f$ value of $1/.130 = 7.69$ $z$ units, while that for Arnoult's data has an $f$ value of $1/.358 = 2.79$ $z$ units. A relatively simple experimental resolution of the issue would be possible by extending the range of $n$ values from 0 out to frequencies of 50 or 100. If Equation 1 is the correct form of the familiarity-frequency law, very little further rise should occur in the two functions as $n$ in-

[1] This work was supported in part by a contract between the office of Naval Research and Montana State University. R. W. Chambers and Q. A. Leland checked the statistical analyses.
Reproduction in whole or in part is permitted for any purpose of the United States Government.
[2] Arnoult kindly made available some of his data for reanalysis in our laboratory. His index of .950 (Arnoult, p. 274) for a linearized Equation 1 is erroneous; the value of $p$ applied to data derived by the Garner-Hake method is actually .974 ($N = 10$).

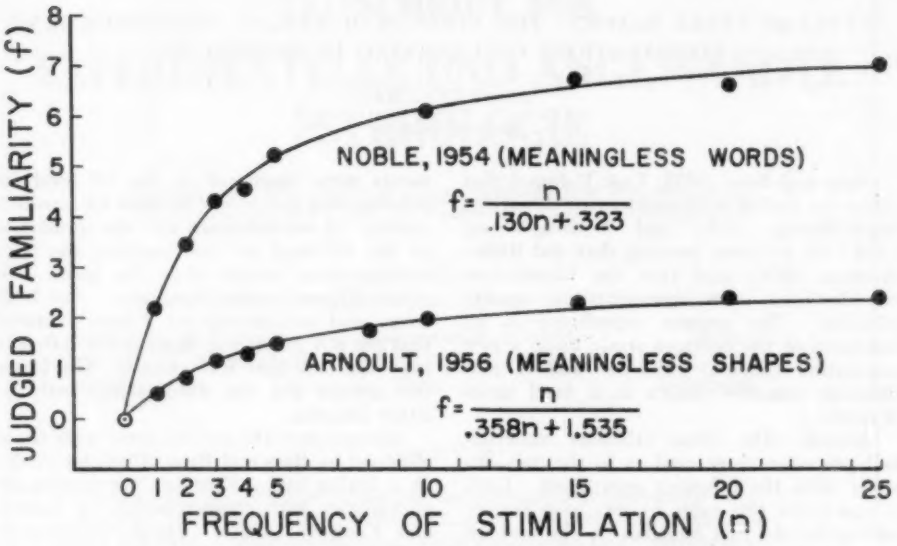



FIG. 1. The familiarity-frequency relationship for two different classes of stimuli.

creases without bound. Since the $f$ scores lie on an interval scale, the 0–25 segment shown in Fig. 1 would have to be re-explored and the same *Ss* should be employed at all points on the curve.

Until the range of observations is increased well beyond $n = 25$, one may continue to entertain an hypothesis (Equation 1) which predicts behavior with less than 1% error, which specifies a rational origin and finite asymptotes, which is more parsimonious than Arnoult's special logarithmic function (Equation 2), and which applies equally well to two classes of stimuli. Referring again to Fig. 1, the writer reaffirms his contention (Noble, 1953) that familiarity is a learnable attribute of stimuli, and that Fig. 1 contains acquisition curves rather than psychophysical functions. This associational hypothesis derives not from hyperbolic-logarithmic distinctions but from a consideration of the fundamental operations involved in the process of familiarization (Noble, 1954, p. 15). The S-R law represented by Equation 1 appears to hold as well for the stimulus class of random shapes (Arnoult, 1956) as for that of meaningless words (Noble, 1954). This is not true of Equation 2, whose goodness-of-fit index is only 94.65% ($N = 9$). In every instance the hyperbolic law predicts the data better than the logarithmic law when familiarity is defined by the original scaling operations (Noble, 1953).

Finally, experimental attempts to delimit the law's invariance by systematically exploring other sensory modalities should be undertaken in conjunction with theoretical studies designed to clarify the basic learning process by means of which frequency exerts its effect on familiarity.

## REFERENCES

ARNOULT, M. D. Familiarity and recognition of nonsense shapes. *J. exp. Psychol.*, 1956, 51, 269–276.

NOBLE, C. E. The meaning-familiarity relationship. *Psychol. Rev.*, 1953, 60, 89–98.

NOBLE, C. E. The familiarity-frequency relationship. *J. exp. Psychol.*, 1954, 47, 13–16.

(Received August 3, 1959)

*Journal of Experimental Psychology*
Vol. 59, No. 6, 1960

# SUPPLEMENTARY REPORT: THE EFFECTS OF VERBAL REINFORCEMENT COMBINATIONS ON LEARNING IN CHILDREN[1]

CAROLYN CURRY

*Indiana University*

Buss and Buss (1956, Exp. I) found that using the verbal reinforcement combinations Right-Wrong (RW) and Nothing-Wrong (NW) led to faster learning than did Right-Nothing (RN), and that the former two combinations were approximately equally effective. The present experiment is an extension of the previous study using a new population (normal children) and a new response measure (errors in a fixed series of trials).

*Method.*—The same stimulus materials and procedure were used as in the previous study with the following exceptions. Each *S* was given 150 cards to sort with correct sorting for the first 50 cards on the basis of color and for the next 100 cards on the basis of number. Ninety grade school pupils were assigned to three groups of boys and three groups of girls. Each group was matched for IQ level on the basis of the Otis Quick Scoring Test of Mental Ability, and each group contained identical proportions of Grades 5 and 6. One group of each sex was assigned to each of the three verbal reinforcement combinations.

*Results.*—The means and *SD*s of the total number of errors made by the three treatment groups during each sorting are shown in Table 1. These data were subjected to analyses of variance (Treatments × IQ level × Sex) with each basis for sorting being considered separately. The effects of Treatments were significant at the .01 level for color-sorting and at the .05 level for number-sorting. A sex difference was also significant at the .05 level for color-sorting, the boys making fewer errors than the girls. No other differences were significant. For both color- and number-sorting, *t* tests revealed that the RN group was significantly different from the NW and RW groups. The latter two groups did not differ significantly in either instance.

TABLE 1

MEANS AND *SD*s OF ERRORS FOR VERBAL REINFORCEMENT COMBINATION GROUPS

| Group | Color-sorting | | Number-sorting | |
|---|---|---|---|---|
| | Mean | *SD* | Mean | *SD* |
| RW | 14.83 | 17.45 | 28.83 | 32.97 |
| RN | 26.87 | 23.00 | 44.90 | 43.93 |
| NW | 14.16 | 17.16 | 27.97 | 32.68 |

*Discussion.*—The results agree with those obtained by Buss and Buss (1956) for adults on a similar task. However, the results are at variance with those reported by Terrell and Kennedy (1957). These investigators studied the effects of a number of types of reinforcement on learning in children. Among these were *E*'s verbalizations to indicate only correct responses for one group and only incorrect responses for a second group. Although these two groups are somewhat comparable to the RN and NW groups of the present experiment, there were no significant differences in the effectiveness of these two combinations. However, the Terrell and Kennedy procedure also included a light flash which indicated correct responses to all *S*s. In the present experiment, and in that reported by Buss and Buss, the *S*'s only source of information about the correctness of a response was the assigned verbal reinforcement combination. The difference between these results and those obtained by Terrell and Kennedy may be a result of the presence of additional information provided to *S*s about their responses in the latter experiment. It seems possible that the differential effectiveness of the verbal reinforcement combinations may be dependent on there being no other information about the correctness of a response available.

[1] This paper is based on an M.A. thesis submitted to the graduate school of Indiana University. The author wishes to express her appreciation to A. M. Buchwald, thesis chairman, for his many helpful suggestions. The major results of this study were presented at the 1959 meeting of the APA.

## REFERENCES

BUSS, A. H., & BUSS, E. H. The effect of verbal reinforcement combinations on conceptual learning. *J. exp. Psychol.*, 1956, **52**, 283–287.

TERRELL, G., JR., & KENNEDY, W. A. The discrimination learning and transposition in children as a function of the nature of the reward. *J. exp. Psychol.*, 1957, **53**, 257–260.

(Received November 19, 1959)

# ZEITSCHRIFT FÜR
# EXPERIMENTELLE UND ANGEWANDTE PSYCHOLOGIE

## Organ der Deutschen Gesellschaft für Psychologie


Herausgeber für Deutschland
PROF. DR. PH. LERSCH, München; PROF. DR. F. SANDER, Bonn

| Herausgeber für Österreich | Herausgeber für die Schweiz | Redaktion |
|---|---|---|
| PROF. DR. H. ROHRACHER, Wien | PROF. DR. R. MEILI, Bern | DR. C. J. HOGREFE, Göttingen |

Band VI/3



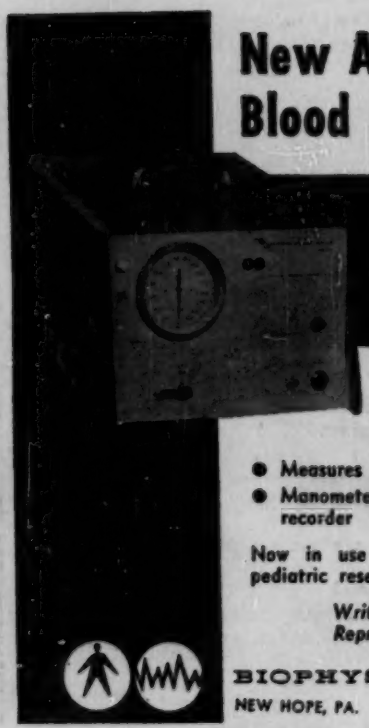
CARL SCHNEIDER. Zum Gestaltbegriff der griechischen Spätantike
JOHANNES THYSSEN. Ontologische Schichtung und Ganzheitspsychologie
HANS VOLKELT. Simultangestalten, Verlaufsgestalten und "Einführung"
ELISABETH LIPPERT. Über motorisch geleistete absolute Distanzschätzungen
FWU TARNG DUN. Some reconsiderations of the integrative action of the nervous system in special reference to the psychology of Gestalt and development
CARL-FRIEDRICH GRAUMANN. Aktualgenese. Die deskriptiven Grundlagen und theoretischen Wandlungen des aktualgenetischen Forschungsansatzes
JOHANNES LINSCHOTEN. Aktualgenese und heuristisches Prinzip
WOLFGANG BÖHM. Methodische Möglichkeiten der experimentellen Wahrnehmungserschwerung bei der auditiven Aktualgenese
JOHANNES VOIGT. Die Aktualgenese im Denkprozeß
WILHELM WITTE. Aktualgenese der Erinnerung
THEODOR SCHARMANN. Aktualgenetisches zum Rorschachtest
JOSEF DERBOLAV. Anamnesis und Aktualgenese. Zur Genealogie von Gestalt und Ganzheit
WILHELM JOSEF REVERS. Über das Wollen. Genese des Wollens: Abhebung vom Fühlen
KARLFRIED GRAF DÜRCKHEIM. Inbild und Gestalt
UDO UNDEUTSCH. Neuere Untersuchungen zur Altersgestalt der Pubeszenz
FRIEDRICH WINNEFELD. Gestaltauffassung und Umgestaltung in genetischer Sicht
ARNULF RÜSSEL. Die Auslöser der Instinkthandlungen
LEO WEISGERBER. Energetische Terminologie in der Sprachpsychologie
GERHARD FUNKE. Bewußtsein und Gewissen kulturtheoretisch
WILHELM SALBER. Sind Ganzheiten praktisch?
CARL FERVERS. Zur Psychologie des Filmerlebens
ERICH FELDMANN. Entwicklungspsychologie und Erziehungswissenschaft
HANS MIESKES. Formen, Phasen und Fundamentalstufen der Bildung
GOTTFRIED HAUSMANN. Ganzheit und Aktualgenese in ihrer Bedeutung für die Methodik des Unterrichts
ALBERT WELLEK. Das Prägnanzproblem der Gestaltpsychologie und das "Exemplarische" in der Pädagogik



All articles are provided with summaries in English and French.

The journal appears quarterly and contains 160 pages per number.
Subscription is DM 64.—per Volume

VERLAG FÜR PSYCHOLOGIE—DR. C. J. HOGREFE—GÖTTINGEN
Brentanoweg 10, Germany



# New Advance in Blood Pressure Measurement

## CONTINUOUS SYSTOLIC MONITOR

- Accurate *continuous* measurement of blood pressure
- No arterial puncture—adult, pediatric and animal application
- Measures response to emotional and external stimuli
- Manometer pressure gauge plus connection for auxiliary recorder

Now in use in pharmacology laboratories, psychological and pediatric research, operating and recovery rooms.

*Write for Brochure CSM-4 or Demonstration Representatives throughout U.S. and Canada*

BIOPHYSICAL ELECTRONICS, INC.
NEW HOPE, PA.